Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **491,** 1190–1212 (2020)
Advance Access publication 2019 November 05

doi:10.1093/mnras/stz3011

Cosmic rays or turbulence can suppress cooling flows (where thermal heating or momentum injection fail)

Kung-Yi Su,[1,2]★ Philip F. Hopkins,[1] Christopher C. Hayward,[2,3] Claude-André Faucher-Giguère,[4] Dušan Kereš,[5] Xiangcheng Ma,[1,6] Matthew E. Orr,[1] T. K. Chan[5] and Victor H. Robles[7]

[1]*TAPIR 350-17, California Institute of Technology, 1200 E. California Boulevard, Pasadena, CA 91125, USA*
[2]*Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA*
[3]*Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA*
[4]*Department of Physics and Astronomy and CIERA, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA*
[5]*Department of Physics, Center for Astrophysics and Space Sciences, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA*
[6]*Department of Astronomy and Theoretical Astrophysics Center, University of California Berkeley, Berkeley, CA 94720, USA*
[7]*Center for Cosmology, Department of Physics and Astronomy, University of California, Irvine, CA 92697, USA*

Accepted 2019 October 19. Received 2019 October 3; in original form 2018 December 11

ABSTRACT

The quenching 'maintenance' and 'cooling flow' problems are important from the Milky Way through massive cluster elliptical galaxies. Previous work has shown that some source of energy beyond that from stars and pure magnetohydrodynamic processes is required, perhaps from active galactic nuclei, but even the qualitative form of this energetic input remains uncertain. Different scenarios include thermal 'heating', direct wind or momentum injection, cosmic ray heating or pressure support, or turbulent 'stirring' of the intracluster medium (ICM). We investigate these in 10^{12}–$10^{14}\,M_\odot$ haloes using high-resolution non-cosmological simulations with the FIRE-2 (Feedback In Realistic Environments) stellar feedback model, including simplified toy energy injection models, where we arbitrarily vary the strength, injection scale, and physical form of the energy. We explore which scenarios can quench without violating observational constraints on energetics or ICM gas. We show that turbulent stirring in the central ~ 100 kpc, or cosmic ray injection, can both maintain a stable low-star formation rate halo for >Gyr time-scales with modest energy input, by providing a non-thermal pressure that stably lowers the core density and cooling rates. In both cases, associated thermal-heating processes are negligible. Turbulent stirring preserves cool-core features while mixing condensed core gas into the hotter halo and is by far the most energy efficient model. Pure thermal heating or nuclear isotropic momentum injection require vastly larger energy, are less efficient in lower mass haloes, easily overheat cores, and require fine tuning to avoid driving unphysical temperature gradients or gas expulsion from the halo centre.

Key words: MHD – turbulence – methods: numerical – cosmic rays – galaxies: clusters: intracluster medium – X-rays: galaxies: clusters.

1 INTRODUCTION

How to 'quench' the *massive* galaxies and keep them 'red and dead' over a large fraction of cosmic time, at stellar masses $\gtrsim 10^{11}\,M_\odot$ (above $\sim L_*$ in the galaxy luminosity function), has been a major outstanding problem in galaxy formation for decades (see e.g. Bell et al. 2003; Kauffmann et al. 2003; Madgwick et al. 2003; Baldry et al. 2004; Blanton et al. 2005; Kereš et al. 2005; Dekel & Birnboim 2006; Kereš et al. 2009; Pozzetti et al. 2010; Wetzel, Tinker & Conroy 2012). The major difficulty lies in the classic 'cooling flow' problem – X-ray observations have found significant radiative cooling in the hot gas of elliptical galaxies and clusters, indicating cooling times shorter than a Hubble time (Fabian et al. 1994; Peterson & Fabian 2006). However, comparing the inferred cooling flow (reaching up to $\sim 1000\,M_\odot\,\mathrm{yr}^{-1}$ in clusters), neither sufficient cold gas from H I and CO observations (McDonald, Veilleux & Mushotzky 2011; Werner et al. 2013) nor sufficient star formation (Tamura et al. 2001; O'Dea et al. 2008; Rafferty, McNamara & Nulsen 2008) have been found in galaxies. Simulations and

★ E-mail: kungyisu@gmail.com

semi-analytic models, which do not suppress cooling flow and simply allow gas to cool into the galactic core, typically predict over an order of magnitude higher star formation rates (SFRs) than observed (for recent examples, see e.g. the weak/no feedback runs in Sijacki et al. 2007; Booth & Schaye 2009; Choi et al. 2015; Li et al. 2015).

To compensate for the observed cooling, there must be some sort of heat source or pressure support. Moreover, the heat must still preserve the cool-core (CC) structure in the majority of galaxies according to the observations (Peres et al. 1998; Mittal et al. 2009). One way to achieve this is to suppress the cooling flow and maintain a very low-SFR stable CC cluster. Another possibility is that clusters undergo CC – non-cool-core (NCC) cycles: a stronger episode of feedback overturns the cooling flows, resulting in an NCC cluster, which gradually recovers to a CC cluster and start another episode of feedback.

The various non-active galactic nuclei (AGNs) solutions to the cooling flow problem proposed in the literature generally belong to the former case, as they are mostly steady heating mechanisms. These generally invoke physics that are unambiguously present, but play an uncertain role in quenching and/or the cooling flow problem, including: stellar feedback from shock-heated asymptotic giant branch (AGB) winds (Conroy, van Dokkum & Kravtsov 2015), Type Ia supernovae (SNe; e.g. Sharma et al. 2012, and references therein), or SNe-injected cosmic rays (CRs; Pfrommer et al. 2017; Ruszkowski, Yang & Zweibel 2017a; Butsky & Quinn 2018; Farber et al. 2018; Jacob et al. 2018); magnetic fields (Soker & Sarazin 1990; Beck et al. 1996, 2012) and thermal conduction (Binney & Cowie 1981; Tucker & Rosner 1983; Fabian, Voigt & Morris 2002; Voigt et al. 2002; Zakamska & Narayan 2003) in the circumgalactic medium (CGM) or intracluster medium (ICM); or 'morphological quenching' via altering the galaxy morphology and gravitational stability properties (Dekel, Sari & Ceverino 2009; Martig et al. 2009). Although these processes can slightly suppress the star formation and 'help' suppress the cooling flows, most previous studies, including our own exhaustive survey studying each of these in simulations similar to those presented here (Su et al. 2019, hereafter Paper I), have shown that they do not fundamentally alter the classic cooling flow picture. In the end, the star formation is still cooling flow regulated, and the SFR is orders of magnitude too high.

Consequently, AGN feedback seem to be the most promising possible solution to the cooling flow problem, and there has been a tremendous amount of theoretical work on the topic (for recent studies see Eisenreich et al. 2017; Gaspari & Sądowski 2017; Jacob & Pfrommer 2017a,b; Li, Ruszkowski & Bryan 2017; Weinberger et al. 2017a; Li et al. 2018; Pellegrini et al. 2018; Yoon et al. 2018; Martizzi et al. 2019 and see e.g. Silk & Rees 1998; Fabian 1999; Ciotti & Ostriker 2001; Hopkins et al. 2005, 2006a; Croton et al. 2006; McNamara & Nulsen 2007; Guo & Oh 2008; Ciotti, Ostriker & Proga 2009; Ostriker et al. 2010; Choi et al. 2012; Pfrommer 2013; Wiener, Oh & Guo 2013 for earlier works). Observations show that the available energy budget can easily be comparable to the cooling rate, and unambiguous cases of AGN expelling gas from galaxies, injecting thermal energy via shocks or sound waves or photoionization and Compton heating, 'stirring' the CGM and ICM, and creating 'bubbles' of hot plasma with non-negligible relativistic components, are ubiquitous (see e.g. Hickox & Alexander 2018, for a detailed review).

However, despite its plausibility and the extensive work above, the detailed physics of AGN feedback remains uncertain, as do the relevant 'input parameters'. Unlike stellar feedback, where we have strong theoretical and observational constraints on SNe event rates, energy inputs, metal yields, etc., AGN properties like energetics, kinetic luminosities, duty cycles, geometries, and their dependence on the black hole (BH) mass and accretion are much less well constrained. Besides, even with the same energy input rate, how and where the energy is coupled to the CGM and ICM remain highly uncertain.

Therefore, instead of jumping into a specific (potentially more realistic) AGN feedback model, in this study we 'take a step back' and explore various idealized AGN 'toy models' with energy injection in different forms (e.g. direct isotropic momentum injection, turbulent stirring, thermal heating, CR injection), acting on different spatial scales, and with different energetics. Our goal is to answer the following simple questions: (i) What form[s] of energy input (if any) can possibly quench a cooling flow, *without* generating unrealistic galaxy or halo properties in obvious disagreement with observations? For example, one could easily imagine scenarios which 'quench' galaxies by simply expelling all the gas in the halo – but this would violate the wealth of observations indicating massive haloes retain most of the cosmological baryon fraction (e.g. Giodini et al. 2009; Gonzalez et al. 2013; Sanderson et al. 2013) (let alone more detailed constraints on density/temperature/entropy profiles). (ii) If any form of energy injection is viable, over what (order of magnitude) spatial scales must it act? In other words, if the energy is primarily deposited around the galactic nucleus, does this yield behaviour that is 'too explosive'? Does the injection have to be fine tuned to occur where the cooling is occurring? (iii) Likewise, what are the required energetics, and are they reasonable compared to observational constraints and plausible accretion efficiencies of supermassive black holes in these systems? (iv) If a model quenches, what is the actual mechanism? For example, turbulent stirring could suppress cooling flows via heating through thermalized kinetic energy (viscous or shock heating), or through providing non-thermal pressure that 'holds up' the halo despite its cooling, or through bulk mixing of cold and hot gas. (v) Does the model quench by maintaining a low-SFR stable CC cluster or turning it into an NCC cluster? If it is the latter case, how long (if ever) does it take to recover a CC after the injection is turned off?

All of these questions have been studied to varying extent in the literature already (see references above). And we will argue below that our conclusions are largely consistent with these previous works. But this manuscript expands on these previous studies in at least three important ways. (i) We attempt a broader and more comprehensive survey, across a variety of energy injection mechanisms, scales, and energetics, in different halo masses, using an otherwise identical set of physics and numerics, to enable fair comparisons. (ii) We aim to implement all of these in fully 'live', global simulations that self-consistently (and simultaneously) treat the entire halo and star-forming galactic disc. For such global simulations, our survey also reaches higher resolution compared to most previous works, allowing us to resolve more detailed substructure in the CGM and galactic disc. (iii) We include explicit, detailed treatments of radiative cooling, the multiphase interstellar medium (ISM) and CGM, star formation, and stellar feedback following the Feedback In Realistic Environments (FIRE)[1] simulations (Hopkins et al. 2014; Muratov et al. 2015; Hopkins et al. 2018b), in order to more robustly model both the gas dynamics and the response of galactic SFRs to cooling flows.

In Section 2, we summarize the initial conditions and the AGN toy models considered here, and describe our numerical simulations.

[1]FIRE project website: *http://fire.northwestern.edu*

The results for our 10^{14} M$_\odot$ halo are presented in Section 3. In Section 4, we test the toy models in smaller haloes in the mass range 10^{12}–10^{13} M$_\odot$. We discuss the effects of each feedback toy model in turn, and explain why it works or not, in Section 5. We summarize in Section 6. Readers not interested in the full parameter study can jump directly from Section 2 to Section 5, in which we summarize the key properties of the more successful 10^{14} M$_\odot$ runs in Fig. 13.

2 METHODOLOGY

We perform simulations of isolated galaxies with halo masses ranging from 10^{12} to 10^{14} M$_\odot$ to test how various physical processes affect galactic cooling flows. We set-up the initial conditions according to the observed profiles of CC clusters at low redshift, as detailed in Section 2.1. Without proper quenching mechanisms, although the galaxies have initial properties consistent with observations, their cooling flow rates and SFRs quickly run away, surpassing the observational values by orders of magnitude. We evolve the simulations with various energy injection models and test to what extent (if any) they suppress the cooling flow and whether they can maintain stably quenched galaxies. Our intention is not to answer where the energy came from or how the scales arise but rather to ask more simply 'are there any combinations of energy/momentum/turbulence and driving/injection scale that can give potentially viable solutions'?[2]

Our simulations use GIZMO (Hopkins 2015),[3] in its meshless finite mass (MFM) mode, which is a Lagrangian mesh-free Godunov method, capturing advantages of grid-based and smoothed-particle hydrodynamics (SPH) methods. Numerical implementation details and extensive tests are presented in Hopkins (2015).

Our default simulation uses the FIRE-2 implementation of the FIRE physical treatments of the ISM and stellar feedback, the details of which are given in Hopkins et al. (2018a, b), along with extensive numerical tests. Cooling is followed from 10 to 10^{10} K, including the effects of photoelectric and photoionization heating, collisional, Compton, fine structure, recombination, atomic, and molecular cooling. Star formation is treated via a sink particle method, allowed only in molecular, self-shielding, locally self-gravitating (Hopkins, Narayanan & Murray 2013) gas, above a density $n > 100\,\mathrm{cm}^{-3}$. Star particles, once formed, are treated as a single stellar population with metallicity inherited from their parent gas particle at formation. All feedback rates (SNe and mass-loss rates, spectra, etc.) and strengths are initial mass function (IMF)-averaged values calculated from STARBURST99 (Leitherer et al. 1999) with a Kroupa (2002) IMF. The feedback model includes: (1) Radiative feedback including photoionization and photoelectric heating, as well as single- and multiple-scattering radiation pressure tracked in five bands (ionizing, FUV, NUV, optical-NIR, IR). (2) Stellar particles continuously lose mass and inject mass, metals, energy, and momentum in the form of OB and AGB winds. (3) Type II and Ia SNe (including both prompt and delayed populations) happen stochastically according to the tabulated rate. Once they occur, the stellar particles lose mass and inject the appropriate mass, metal, momentum, and energy to the surrounding gas.

2.1 Initial conditions

The initial conditions (ICs) studied here are presented and described in detail in Paper I. Their properties are summarized in Table 1. In this paper, the bulk of our study will initially focus on the **m14** halo, which has the most dramatic (massive) cooling flow (we will then consider the other haloes in turn). The dark matter (DM) halo, bulge, black hole, and gas + stellar disc are initialized following Springel & White (1999) and Springel (2000). We assume a spherical, isotropic, Navarro, Frenk & White (1996) profile DM halo; a Hernquist (1990) profile stellar bulge; an exponential, rotation-supported disc of gas and stars (10^{10} and 2×10^{10} M$_\odot$) initialized with Toomre $Q\approx1$; a BH with mass 1/300 of the bulge mass (e.g. Häring & Rix 2004); an extended spherical, hydrostatic gas halo with a β-profile ($\beta = 1/2$) and rotation at twice the net DM spin (so $\sim$ 10–15 per cent of the support against gravity comes from rotation, the rest thermal pressure resulting from the virial shock). The initial metallicity drops from solar ($Z = 0.02$) to $Z = 0.001$ with radius as $Z = 0.02\,(0.05 + 0.95/(1 + (r/20\,\mathrm{kpc})^{1.5}))$. For the runs with CR injection, initial magnetic fields are azimuthal with $|\boldsymbol{B}| = 0.3\,\mu\mathrm{G}/(1 + (r/20\,\mathrm{kpc})^{0.375})$ (extending throughout the ICM), and initial CR energy density is in equipartition with the local initial magnetic energy density. The ICs are run adiabatically (no cooling or star formation) to relax any initial transients before use.

The ICs are designed to be similar to observed CC systems of similar mass wherever possible (see e.g. Humphrey et al. 2012; Humphrey & Buote 2013; Su, White & Miller 2013; Su et al. 2015). Our **m14** halo has initial cooling rate at $\sim 8\times10^{43}$ erg s^{-1}, with $\sim 3\times10^{43}$ erg s^{-1} radiated in X-ray (0.5–7 kev).

In **m12** and **m13** the mass resolution is constant; in **m14** (given its much larger total mass but the need to ensure fixed physical mass resolution in e.g. the star-forming disc) the resolution here matches run 'MR-MRS' in Paper I, adopting a radially dependent super-Lagrangian refinement scheme. The target gas mass resolution is set to $= 3\times10^4$ M$_\odot$ inside $r < 10$ kpc, and increases smoothly $\propto r$ outside this radius up to a maximum $= 2\times10^6$ M$_\odot$ at $\sim$ 300 kpc. Gas resolution elements are automatically merged or split appropriately if they move inward/outward, to maintain this mass resolution (to within a factor =2 tolerance) at all times. A resolution study is included in the appendix of Paper I.

2.2 Energy injection models surveyed

The toy models we investigate include momentum injection (simulations prefixed 'Momm'), turbulent stirring ('Turb'), thermal input ('Th'), and CR input ('CR'). All the simulations are listed in Table 2, which also tabulate the energy and momentum input within different ranges. The 'Default' run includes only 'FIRE-2' stellar feedback. The other runs have various AGN toy models implemented on top of 'FIRE-2' stellar feedback. Only the runs with CR injection have magnetic fields. The runs labelled as 'BH' have energy (momentum) injected in the black hole neighbourhood, while the 'core' runs have a wider distributed injection with the kernel functions listed in the last column of Table 2. The other runs labelled 'uni' have uniform input per unit gas mass (so most of the energy is deposited at large radii). The detailed radial dependence of the energy and momentum input is shown in Fig. 1. The simulation duration is also listed in

[2] We note that while there are more constraints from X-ray observations for rich clusters of mass $\sim 10^{15}$ M$_\odot$, we focus on the mass range of 10^{12}–10^{14} M$_\odot$ both for reasons of computational expense and because 'quenching', 'overcooling', and variants of the classical cooling-flow problem appear in simulations absent AGN feedback as soon as L* is reached. The same parameters would not necessarily work for a rich cluster – hence, we repeat some of our parameter survey at each halo mass to obtain a sense of the scaling.

[3] A public version of this code is available at http://www.tapir.caltech.edu/~phopkins/Site/GIZMO.html.

Table 1. Properties of initial conditions for the simulations/haloes studied here.

Model	Resolution		DM halo			Stellar bulge			Stellar disc		Gas disc		Gas halo	
	ϵ_g (pc)	m_g ($M_\odot$)	M_{halo} ($M_\odot$)	r_{dh} (kpc)	V_{Max} (km s^{-1})	M_{bar} ($M_\odot$)	M_b ($M_\odot$)	a (kpc)	M_d ($M_\odot$)	r_d (kpc)	M_{gd} ($M_\odot$)	r_{gd} (kpc)	M_{gh} ($M_\odot$)	r_{gh} (kpc)
m12	1	8e3	1.5e12	25	174	2.2e11	1.5e10	1.0	5.0e10	3.0	5.0e9	6.0	1.5e11	25
m13	3	5e4	1.0e13	100	240	7.2e11	1.0e11	2.8	1.4e10	2.8	5.0e9	2.8	6.0e11	10
m14	**1**	**3e4**	**8.5e13**	**220**	**600**	**1.5e13**	**2.0e11**	**3.9**	**2.0e10**	**3.9**	**1e10**	**3.9**	**1.5e13**	**22**

Note. Parameters of the galaxy models studied here (Section 2.1): (1) Model name. The number following 'm' labels the approximate logarithmic halo mass. (2) ϵ_g: Minimum gravitational force softening for gas (the softening for gas in all simulations is adaptive, and matched to the hydrodynamic resolution; here, we quote the minimum Plummer equivalent softening). (3) m_g: Gas mass (resolution element). There is a resolution gradient for m14, so its m_g is the mass of the highest resolution elements. (4) M_{halo}: Halo mass. (5) r_{dh}: NFW halo scale radius (the corresponding concentration of m12, m13, m14 is $c = 12$, 6, 5.5, respectively). (6) V_{max}: Halo maximum circular velocity. (7) M_{bar}: Total baryonic mass. (8) M_b: Bulge mass. (9) a: Bulge Hernquist-profile scale length. (10) M_d : Stellar disc mass. (11) r_d : Stellar disc exponential scale length. (12) M_{gd}: Gas disc mass. (13) r_{gd}: Gas disc exponential scale length. (14) M_{gh}: Hydrostatic gas halo mass. (15) r_{gh}: Hydrostatic gas halo $\beta = 1/2$ profile scale length.

Table 2. Physics variations (run at highest resolution) in our halo-**m14** survey.

Model	Summary	ΔT (Gyr)	$\dot{E}_{tot}$	$\dot{E}_{r<30}$	$\dot{E}_{r<100}$	$\dot{E}_{r>100}$ (erg s^{-1})	$\dot{P}_{tot}$	$\dot{P}_{r<30}$	$\dot{P}_{r<100}$	$\dot{P}_{r>100}$ (g cm s^{-2})	Kernel (r in kpc)
Default	–	2.0	–	–	–	–	–	–	–	–	–
Momm-BH-34	Minor	0.4	4.7–7.6 e40	–	–	–	1.2e34	–	–	–	BH neighbour
Momm-BH-35	Minor	2.0	1.8–3.6 e42	–	–	–	1.2e35	–	–	–	BH neighbour
Momm-BH-36	Explosive	0.3	2.0–0.07 e44	–	–	–	3.6e36	–	–	–	BH neighbour
Turb-uni-1	Minor	2.0	3.6–7.7 e42	5.2–17 e39	4.6–5.7 e40	3.5–7.6 e42	1.6–1.6 e35	3.3–5.1 e32	3.0–1.7 e33	1.6–1.6 e35	Uniform
Turb-uni-2	$L_X\downarrow$	2.0	6.5–20 e42	9.5–11 e39	8.4–5.8 e40	6.4–20 e42	2.9–2.9 e35	6.0–3.4 e32	5.5–1.5 e33	2.9–2.9 e35	Uniform
Turb-core-1	Moderate	2.0	5.0–5.6 e41	8.6–15 e39	6.2–8.5 e40	4.4–4.8 e41	2.5–2.0 e34	5.5–4.5 e32	4.1–2.3 e33	2.1–1.8 e34	$a \sim \exp(-r/200)$
Turb-core-2	Moderate	2.0	1.6–2.4 e41	9.1–18 e39	6.7–11 e40	9.8–14 e40	9.8–7.8 e33	5.9–5.4 e32	4.5–2.9 e33	5.3–5.0 e33	$a \sim \exp(-(r/140)^2)$
Turb-core-3	Moderate	2.0	2.1–2.9 e41	1.6–2.1 e40	8.1–12 e40	1.3–1.6 e41	1.2–0.9 e34	9.8–6.1e32	5.4–3.1e33	6.6–6.2 e33	$a \sim 2\exp(-r/80)$
Turb-core-4	**Quenched**	**2.0**	**5.7–5.9 e41**	**2.7–0.5 e40**	**1.4–1.1 e41**	**4.4–4.8 e41**	**3.1–2.0 e34**	**1.7–0.2 e33**	**9.2–2.4 e33**	**2.1–1.8 e34**	$\mathbf{a_{r<100} \sim 3\exp(-(r/79)^2)}$ $\mathbf{a_{r>100} \sim \exp(-r/200)}$
Turb-core-5	Quenched (NCC)	2.0	6.6–6.6 e41	5.2–1.7 e40	2.2–1.8 e41	4.4–4.8 e41	3.6–2.1 e34	3.3–0.3 e33	1.5–0.3 e34	2.1–1.7 e34	$a_{r<100} \sim 6\exp(-(r/66)^2)$ $a_{r>100} \sim \exp(-r/200)$
Turb-core-6	Mach↑	0.7	9.9–11 e41	1.7–0.7 e41	5.5–0.5 e41	4.4–5.7 e41	5.9–2.6 e34	1.1–0.06 e34	3.7–0.6 e34	2.1–2.0 e34	$a_{r<100} \sim 20\exp(-(r/54)^2)$ $a_{r>100} \sim \exp(-r/200)$
Th-uni-43	Minor	1.4	2.1e43	4.2–8.1 e40	3.9–4.0e41	2.0e43	–	–	–	–	Uniform
Th-uni-44	Minor	2.0	2.1e44	4.2–7.3 e41	3.9–3.4 e42	2.0e44	–	–	–	–	Uniform
Th-core-43	Minor	2.0	2.0–1.8 e43	1.2–1.2e43	2.0–1.8e43	6.8–7.2 e38	–	–	–	–	$\dot{E} \propto \exp(-(r/30)^2)$
Th-core-44	**Quenched**	**2.0**	**2.0–0.5 e44**	**1.2–0.3 e44**	**2.0–0.5 e44**	**6.8–6.4 e39**	–	–	–	–	$\mathbf{\dot{E} \propto \exp(-(r/30)^2)}$
Th-core-45	Explosive	1.0	1.9–0.1 e45	1.2–0.02 e45	1.9–0.1 e45	6.8–5.2 e40	–	–	–	–	$\dot{E} \propto \exp(-(r/30)^2)$
Th-BH-43	Minor	2.0	2.1e43	–	–	–	–	–	–	–	BH neighbour
Th-BH-44	Explosive	1.2	2.1e44	–	–	–	–	–	–	–	BH neighbour
Th-BH-45	Explosive	0.4	2.1e45	–	–	–	–	–	–	–	BH neighbour
CR-BH-42	Minor	2.0	2.1e42	–	–	–	–	–	–	–	BH neighbour
CR-BH-43	**Quenched**	**2.0**	**2.1e43**	–	–	–	–	–	–	–	**BH neighbour**
CR-BH-44	Explosive	0.3	2.1e44	–	–	–	–	–	–	–	BH neighbour

Note. This is a partial list of simulations studied here: each was run using halo **m14**, systematically varying the energy injection mechanism, scale, and energetics, at our highest resolution (a broader low-resolution parameter survey, and our survey of haloes **m12** and **m13**, are not included here). Columns list: (1) Model name. Models labelled 'Momm', 'Turb', 'Th', and 'CR' correspond to (radial) momentum injection, turbulent injection or 'stirring', thermal energy injection ('heating'), and cosmic ray (CR) injection, respectively. Models labelled 'uni', 'core', 'BH' adopt different kernels (see Fig. 1). (2) Summary of the results. Minor, moderate, and quenched correspond, respectively, to an SFR of $\gtrsim 10$, ~ 1 to 10, and $\lesssim 1\,M_\odot\,yr^{-1}$. The runs labelled otherwise are quenched while having a major drawback (as labelled). (3) ΔT: Simulation duration. All runs are run to 2 Gyr, unless either the halo is completely 'blown out' or completely unaffected. (4) $\dot{E}_{tot}$, $\dot{E}_{r<30}$, $\dot{E}_{r<100}$, and $\dot{E}_{r>100}$ tabulate the total energy input of the corresponding spherical region (with the two values corresponding to the beginning and end of the run). The energy input of 'Momm' and 'Turb' runs is the energy used to accelerate gas (e.g. difference in kinetic energy) in each time-step. (5) $\dot{P}_{tot}$, $\dot{P}_{r<30}$, $\dot{P}_{r<100}$, and $\dot{P}_{r>100}$ tabulate the momentum input in the corresponding region. (6) kernel: the form of the injection kernel.

Table 2. All runs are run to 2 Gyr, unless either the halo is completely 'blown out' or completely unaffected.

Although we will treat the energy/momentum injection rates as essentially arbitrary in our survey, for context it is worth noting that for an $\sim 10^9\,M_\odot$ BH (about as massive as we expect in our **m14** halo) the Eddington limit is $\sim 10^{47}$ erg s^{-1}. The associated photon momentum flux is $L/c \sim 4 \times 10^{36}$ g cm s^{-2}. For more typical low-luminosity AGNs observed in massive galaxies, the energies associated with e.g. their jets reach $\sim 10^{44}$–10^{45} erg s^{-1} (see Fabian 2012).

2.2.1 *Thermal input ('pure heating')*

Any process that ultimately transfers some energy to gas thermal energy can be said to have a 'heating' component. This can occur via radiative (photoionization, Compton), mechanical (shocked winds/jets, compression), viscous (damped sound waves or turbulence), CR (collisions, streaming instabilities), and other processes. Many models in the literature have invoked the idea that heating from AGNs can effectively offset cooling and drive strong pressure-driven outflows, if roughly a few per cent of the luminosity associated with near-Eddington phases can couple thermally (Begelman 2004; Di Matteo, Springel & Hernquist 2005; Springel, Di Matteo & Hernquist 2005; Hopkins et al. 2006b, c, 2007, 2008; Johansson, Naab & Burkert 2009; Hopkins & Elvis 2010; Ostriker et al. 2010; Faucher-Giguère & Quataert 2012; Dubois et al. 2013; Barai et al. 2014; Weinberger et al. 2017b; Pillepich et al. 2018; Richings & Faucher-Giguère 2018a,b).

To mimic this in an intentionally idealized and simplified manner, we directly add (to the usual self-consistent heating

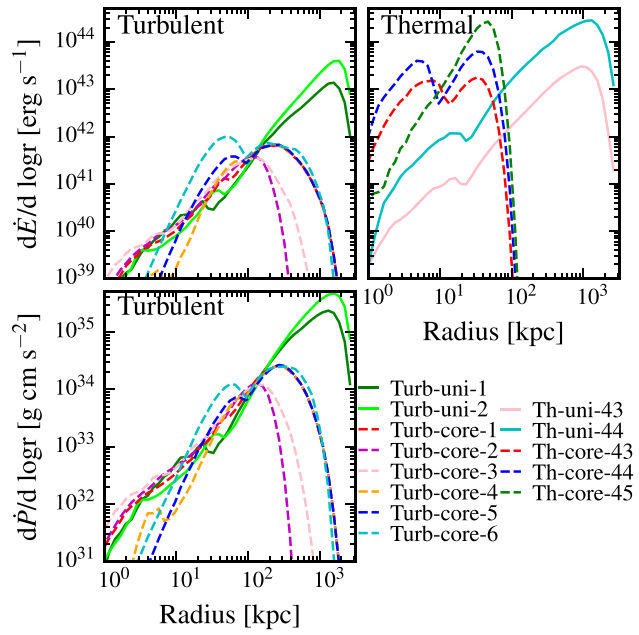

Figure 1. Energy (*top*) or momentum (*bottom*) input rate per unit logarithmic galactocentric radius $\log r$ (time averaged over the last 100 Myr of each run), in a subset of our halo **m14** runs. Runs labelled 'uni' inject these quantities uniformly per unit mass over the whole halo, so large radii (containing most mass/volume) receive most of the injection. Runs labelled 'core' have injection in a Gaussian-like kernel, so most of the energy/momentum ends up around the kernel scale radius. Runs labelled 'BH' inject everything in a kernel centred in the resolution elements immediately surrounding the BH ($\ll$ kpc, hence not shown).

and cooling routines) an analytic heating rate per unit mass $\dot{e}_{\rm inj}(r) = \dot{E}_{\rm tot}\, M_0^{-1}\, f(r)$, where $f(r)$ is a dimensionless spherically symmetric kernel function (centred on the BH at the galaxy centre) normalized to $M_0^{-1} \int \rho(\boldsymbol{x})\, f(|\boldsymbol{x}|)\, d^3\boldsymbol{x} = 1$. We vary both $\dot{E}_{\rm tot}$ and $f(r)$ systematically, as shown in Table 2. In runs labelled 'BH', $f(r)$ is a cubic spline with radius of compact support enclosing the nearest $\sim$96 gas elements to the BH. In runs labelled 'core', $f(r)$ is a Gaussian ($\propto \exp[-(r/r_0)^2]$) with dispersion approximately equal to the β-profile scale length (which is also approximately the critical cooling radius). And in runs labelled 'uni', $f(r)$ is constant out to approximately the virial radius. In the 'BH' cases, $f(r)$ is updated at each time-step, while in the 'core' and 'uni' cases, $f(r)$ is set at the beginning of the runs and kept constant.[4]

2.2.2 *Momentum input*

Again many processes can transfer momentum/kinetic energy to gas, including radiation pressure, mechanical feedback from AGN winds and jets, and 'PdV' work from CR pressure gradients. Again many models have invoked kinetic feedback to suppress cooling flows and SFRs in massive haloes (Gaspari et al. 2011; Choi et al. 2012, 2015; Li et al. 2015; Martizzi et al. 2019) and many have argued it specifically provides a better match to observational constraints and is more efficient compared to 'pure heating' models, especially in the context of 'maintenance' or 'radio mode' feedback (Begelman 2004; Fabian 2012; Dubois et al. 2013; Barai et al. 2014; Meece, Voit & O'Shea 2017; Weinberger et al. 2017b; Pillepich et al. 2018; Martizzi et al. 2019).

Since we will distinguish 'random' or 'non-oriented' driving below, we use this term to refer specifically to models with kinetic feedback oriented strictly radially away from the BH. Moreover because the coupling in the models above is primarily local (and we are not interested for this model in e.g. the case of CRs or hot thermally pressurized gas driving outflows on large scales, since these should be resolved in our 'Thermal Input' and 'CR' runs), we will primarily focus on just the 'BH' (local-kernel) models in this case. In that case a constant momentum flux $\dot{P}$ (directed radially away from the BH) is injected in a similar kernel-weighted fashion among neighbouring gas around the BH (as for thermal energy), but with the kernel weights proportional to the solid angle subtended by each gas element (as seen by the BH).[5]

2.2.3 *Turbulent driving or 'stirring'*

Rather than simply 'pushing outwards', a variety of processes can instead transfer energy to kinetic energy of bulk quasi-random motion, what we call 'turbulent stirring'. AGN bubbles may generate turbulence through Rayleigh–Taylor (RT) and Richtmyer–Meshkov (RM) instabilities (Dimonte & Tipton 2006; Scannapieco & Brüggen 2008; Brüggen & Scannapieco 2009); jets (precessing or not) can drive turbulence through changing bulk motion or secondary instabilities (e.g. Li & Bryan 2014; Yang & Reynolds 2016; Bourne & Sijacki 2017; Martizzi et al. 2019) with driving scale $\sim$ 100 kpc (ZuHone, Markevitch & Zhuravleva 2016; Hitomi Collaboration 2018); and non-AGN processes like halo mergers (e.g. Roettiger, Burns & Loken 1993; Roettiger, Loken & Burns 1997; Norman & Bryan 1999; Ricker & Sarazin 2001; Mitchell et al. 2009; Paul et al. 2011; Vazza et al. 2011), sloshing of cold fronts (e.g. Fujita, Matsumoto & Wada 2004; ZuHone et al. 2013, 2018), and winds from satellites can do likewise (Anglés-Alcázar et al. 2017). Studies have argued turbulence could suppress cooling flows by providing direct pressure support to gas (Parrish et al. 2012), or heating the gas 'directly' via viscous dissipation (Banerjee & Sharma 2014; Zhuravleva et al. 2014), effectively conducting heat from the outer hot halo to the inner CC (Banerjee & Sharma 2014), or mixing cold structures back into hot gas in a thermally unstable medium and thereby efficiently redistributing heat (e.g. Kim & Narayan 2003; Voigt & Fabian 2004; Vernaleo & Reynolds 2006; Parrish, Quataert & Sharma 2010; Ruszkowski & Oh 2010; Banerjee & Sharma 2014).

We represent 'turbulent stirring' by driving turbulence directly following the 'turbulent box' simulations in Bauer & Springel (2012). Turbulence is driven in Fourier space as an Ornstein–Uhlenbeck process (see Schmidt et al. 2009; Federrath et al. 2010; Price & Federrath 2010) with characteristic driving wavelength ($\lambda = 2\pi/k$) set to 1/2 of the halo scale radius (experimenting with this, compared to the kernel or total energy, makes little difference to our conclusions). The compressive part of the acceleration is projected out via a Helmholtz decomposition in Fourier space so that the driving is purely incompressible (solenoidal). After Fourier-transforming back to real space, the stirring is applied as a continuous acceleration $\boldsymbol{a}(\boldsymbol{x})$ to each element; at this stage, we apply the desired kernel function $\boldsymbol{a}(\boldsymbol{x}) \rightarrow \boldsymbol{a}(\boldsymbol{x})\, f(r)\, V_0^{-1}$ (with

[4]This causes the evolution of energy input, especially in the more explosive runs, since the density profiles also evolve.

[5]We emphasize that while this is launched at the BH, it is not a jet model. The scaling with solid angle simply ensures that momentum is launched uniformly in all directions.

$V_0^{-1} \int f(|\boldsymbol{x}|)\, d^3\boldsymbol{x} = 1$). In runs labelled 'uni', $f(r)$ is constant out to approximately the virial radius. In the runs labelled 'core', $f(r)$ is either a Gaussian function or an exponential function as shown in Table 2. The energy and momentum input rates labelled in Table 2 are calculated through $\dot{E} \sim \int dm\ \max(|\boldsymbol{a}(\boldsymbol{x})|)|\boldsymbol{v}|$ and $\dot{P} \sim \int dm\ \max(|\boldsymbol{a}(\boldsymbol{x})|)$, which estimate the upperbounds.[6] We argue that this turbulent stirring is more than an 'effective conductivity' treatment, since the kinetic bulk motion of cold gas elements is important; see Section 5.3 for details.

2.2.4 *Cosmic ray injection*

CRs arise generically from processes that result in fast shocks, so could come from shocked winds or outflows, but are particularly associated with relativistic jets from AGNs (where they can make up the bulk of the jet energy; Berezinsky, Gazizov & Grigorieva 2006; Ruszkowski, Yang & Reynolds 2017b) and hot, relativistic plasma-filled 'bubbles' or 'cavities' (perhaps inflated by jets in the first place) around AGNs. Different authors have argued that they could help suppress cooling flows via providing additional pressure support to gas, driving pressure-driven outflows in the galaxy or CGM, or via heating the CGM/ICM directly via collisional (hadronic and Coulomb) and streaming-instability losses (Guo & Oh 2008; Sharma, Parrish & Quataert 2010; Enßlin et al. 2011; Fujita & Ohira 2011; Fujita, Kimura & Ohira 2013; Pfrommer 2013; Wiener et al. 2013; Jacob & Pfrommer 2017a,b; Pfrommer et al. 2017; Ruszkowski et al. 2017a,b; Jacob et al. 2018)

We treat this analogous to our 'thermal heating' runs – simply injecting CR energy at some fixed rate within a kernel. The CR physics and numerical implementation are described in detail in Chan et al. (2019). Briefly, this treats CRs including streaming (at the local Alfv´en speed, with the appropriate streaming loss term, which thermalizes, following Uhlig et al. 2012, but with $v_{\rm st} = v_{\rm A}$), diffusion (with a fixed diffusivity $\kappa_{\rm cr}$), adiabatic energy exchange with the gas and CR pressure in the gas equation of motion, and hadronic and Coulomb losses (following Guo & Oh 2008). We follow a single energy bin (i.e. GeV CRs, which dominate the pressure), treated in the ultrarelativistic limit. Streaming and diffusion are fully anisotropic along magnetic field lines. In Chan et al. (2019), we show that matching observed gamma-ray luminosities, in simulations with the physics above requires $\kappa_{\rm cr} \sim 10^{29}\,{\rm cm^2\,s^{-1}}$, in good agreement with detailed CR transport models that include an extended gaseous halo around the Galaxy (see e.g. Strong & Moskalenko 1998; Strong et al. 2010; Trotta et al. 2011), so we adopt this as our fiducial value.[7] In practice, because of the large diffusivity, the CR energy density rapidly converges to the same quasi-equilibrium profile regardless of the shape of the injection kernel, so long as the injection scale is not extremely large ($\lesssim 100$ kpc), so we simplify by focusing on the 'BH' kernel choice and keeping the injection isotropic.[8]

[6] Although the acceleration of the gas (as a function of space) is constant in time, the density profiles change. Therefore, the total energy input rates also vary as a function of time.

[7] We caveat that we do not account for the possibility of different diffusion coefficient in different environments.

[8] We also note that, in the runs including CR heating, CRs from SNe are not included, so we have a clean test on the black hole CR injection. We showed in Paper I that CRs from SNe contribute negligibly to quenching, and we note below that the total energy injection from SNe is a factor $\sim 10^2$–10^4 below the analytically input CR energy injection rate.

3 RESULTS IN OUR MASSIVE HALO (M14) SURVEY

As will be shown in the following subsections, 'Th-core-44' ($\dot{E}_{\rm th} \sim \dot{E}_{\rm cool}$), 'Turb-core-4' ($\dot{E}_{\rm turb} < 1$ per cent$\dot{E}_{\rm cool}$), and 'CR-BH-43' ($\dot{E}_{\rm CR} \sim 10$ per cent$\dot{E}_{\rm cool}$) are the more successful runs in the corresponding toy model scenario. We therefore highlight these runs in the subsequent plots, while tuning down the contrast of the 'explosive' runs.

3.1 Star formation history

The first row of Fig. 2 plots the baryonic mass (as a function of time) within 30 kpc ($M_{\rm baryon}^{\rm 30\,kpc}$) excluding the pre-existing stars, which characterizes the cooling flow rates. The second, third, and bottom rows show SFRs, SFRs from gas initially sitting outside 25 kpc (SFRs supplied by the cooling flows), and specific star formation rates (sSFRs), averaged in rolling 10 Myr bins, respectively. Momentum injection below $\sim 10^{35}\,{\rm g\,cm\,s^{-2}} \sim 0.03\,L_{\rm Edd}/c$ does not suppress the cooling flow or star formation by much, while an injection above $3 \times 10^{36}\,{\rm g\,cm\,s^{-2}} \sim L_{\rm Edd}/c$ blows everything away within 50 Myr leaving almost no gas within 70 kpc.

With a lower momentum flux (1–$2 \times 10^{34}\,{\rm g\,cm\,s^{-2}}$; non-radial), turbulent stirring can significantly suppress the cooling flows and star formation. When the turbulent energy input within 100 kpc reaches 1.1–$1.4 \times 10^{41}\,{\rm erg\,s^{-1}}$ ('Turb-core-4'), the core baryonic mass is suppressed by a factor of 3–10. For turbulent energy input rates above $\sim 2 \times 10^{41}\,{\rm erg\,s^{-1}}$ ('Turb-core-5,6'), the SF is eventually completely quenched.

Uniform thermal heating has little effect on the SFRs and cooling flows even if input rate reaches $\sim 10^{44}\,{\rm erg\,s^{-1}}$. Black hole thermal injection, on the other hand, undergoes a sharp transition from having little effect to completely quenching the galaxy by blowing everything away (through a Sedov–Taylor explosion), between injection rates 10^{43} and $10^{44}\,{\rm erg\,s^{-1}}$ (10^{-4}–$10^{-3} L_{\rm Edd}$). The transition is milder if the energy is smoothly injected within a Gaussian kernel of 30 kpc, in which case a stable core baryonic mass and low SFR can be maintained by a heating rate of $\sim 10^{44}\,{\rm erg\,s^{-1}}$ ('Th-Core-44'). However, with a similar cooling flow rate (e.g. 'Turb-core-1' and 'Th-core-43'), turbulent stirring suppresses SFR more efficiently (with a lower energy input rate) than core thermal heating.

Unlike thermal heating, CR energy input can maintain a semistable core baryonic mass and suppressed SFR even if all the energy is deposited in the vicinity of the black hole. The SFR and cooling flows are significantly suppressed by an energy input of $10^{43}\,{\rm erg\,s^{-1}}$, less than the rate required for a thermal heating run with Gaussian kernel to quench. However, when the CR input reaches $10^{44}\,{\rm erg\,s^{-1}}$, the resulting dramatic suppression of core baryonic mass becomes similar to what is caused by the 'explosive' BH-kernel thermal heating.

3.2 The resulting halo properties

3.2.1 *Temperature, density, and entropy*

Fig. 3 shows the average density and luminosity-weighted density, temperature, and entropy as a function of radius averaged over the last 100 Myr of the runs. The shaded regions in the second row indicate the observational density profiles (scaled) for CC (blue)

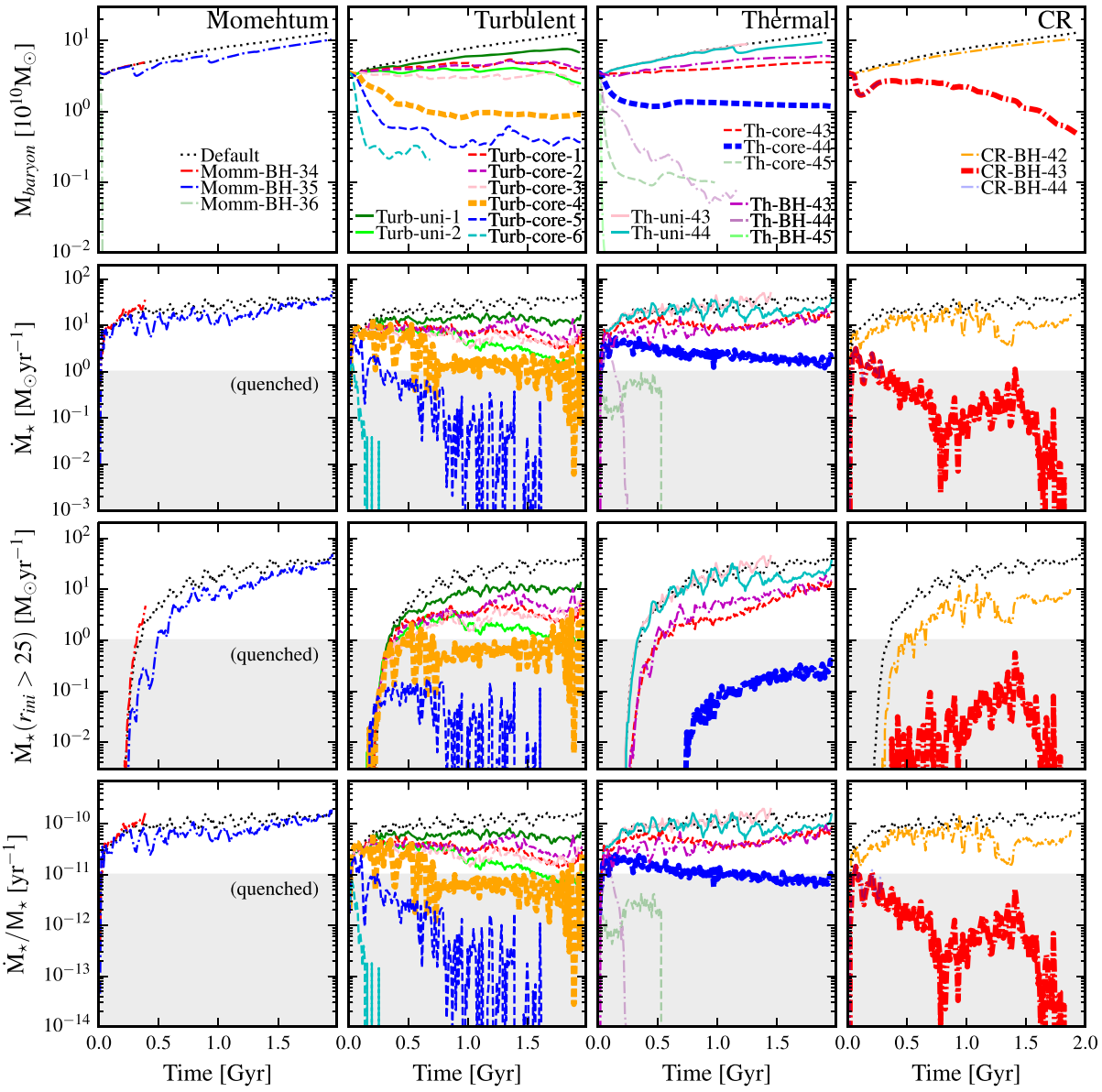

Figure 2. *Top:* Baryonic mass within 30 kpc (excluding pre-existing stars from the ICs), as a function of time, in the halo-**m14** runs from Table 2. This is a proxy for the net amount of cooling-flow inflow. *Second:* SFRs averaged in 10 Myr intervals. *Third:* SFRs specifically from gas which was at $r > 25$ kpc in the ICs (gas which comes in with the cooling flow). *Bottom:* Specific SFRs. The shaded regions indicate the SFR or sSFR that we define as quenched. For each, we compare runs with momentum injection, turbulent stirring, thermal heating, and CR injection (columns, as labelled). Momentum injection below $\dot{P} \lesssim 10^{35}$ g cm s^{-2} does not suppress cooling flows, while $\dot{P} \gtrsim 3 \times 10^{36}$ g cm s^{-2} almost immediately ejects all the gas in the halo. Uniform thermal heating has little effect on SF (most of the energy is 'wasted' at large-r), while nuclear ('BH') injection transitions sharply between doing nothing (the heat is radiated away) and driving a Sedov–Taylor explosion that evacuates the halo around $\dot{E} \sim 3 \times 10^{43}$ erg s^{-1}. Heating with a semi-extended ~ 30 kpc kernel can suppress SF without explosive ejection for $\dot{E}$ carefully chosen around $\dot{E} \sim 10^{44}$ erg s^{-1}. Turbulent stirring more efficiently suppresses SF: when the driving $\dot{E}$ within < 100 kpc reaches $\gtrsim 10^{41}$ erg s^{-1}, the core baryonic mass begins to fall, and by 2× this SF is eventually completely quenched. CR energy input at 10^{43} erg s^{-1} can maintain a low SFR and semistable core baryonic mass even if the energy is deposited in the nucleus.

and NCC (red) clusters (McDonald et al. 2013).[9] The lightened curves in the bottom row indicate the observational entropy profiles for CC (blue) and NCC (red) clusters (McDonald et al. 2013).[10]

[9]We use the panel for $z < 0.1$ in fig. 9 of McDonald et al. (2013) and assume $\rho_{crit} \sim 9.2 \times 10^{30}$ g cm^{-3} and $r_{500} = 650$ kpc (our m14 initial condition).
[10]The haloes in McDonald et al. (2013) have a mass range of $\sim 2 \times 10^{14} < M_{500} < 20 \times 10^{14}$ $M_\odot/h_{70}$. We use their fig. 2 and scale the average entropy at $r = 700$ kpc to 500 kev cm^2 given our halo is smaller (cooler).

Momentum injection does not affect the resulting halo profiles with an input rate less than $\sim 10^{35}$ g cm s^{-2}, while it blows everything away when the input rate reaches $L_{Edd}/c \sim 3 \times 10^{36}$ g cm s^{-2}.

With a lower momentum input at $\sim 10^{34}$ g cm s^{-2}, turbulent stirring can much more efficiently suppress the core density. When the core ($r < 100$ kpc) energy input reaches 1.1–1.4 × 10^{43} erg s^{-1} ('Turb-core-4'), the density suppression becomes more significant. When it reaches 2×10^{43} erg s^{-1} ('Turb-core-5,6'), the core gas is eventually completely heated up, and the entropy profile is flattened. If turbulent stirring is not suppressed at large radii, the

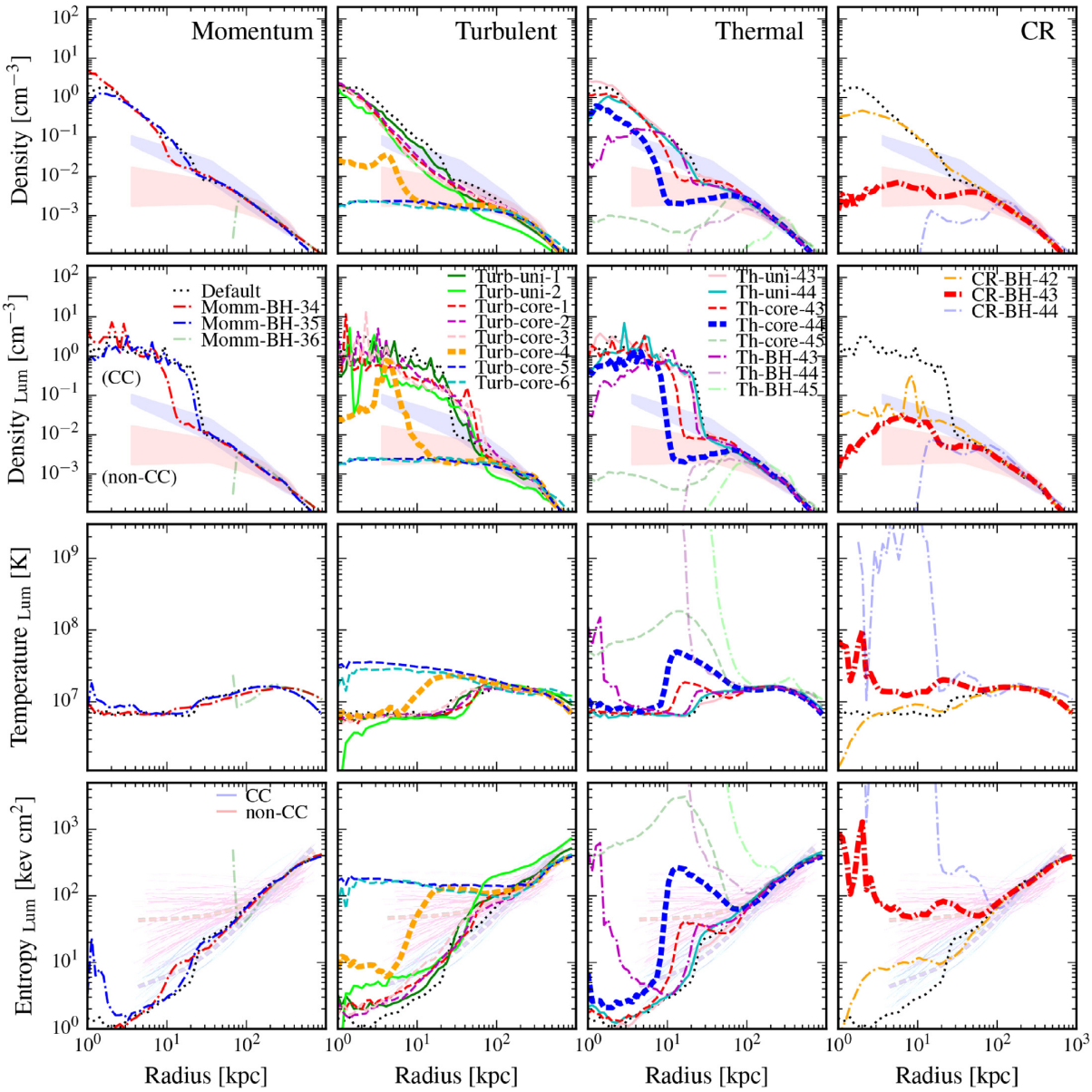

Figure 3. Density (*top*), X-ray cooling luminosity-weighted density (*second*), luminosity-weighted temperature (*third*), and luminosity-weighted entropy (*bottom*) versus radius averaged over the last $\sim$ 100 Myr in the **m14** runs from Fig. 2. The shaded regions in the second row and the lightened curves in the bottom row indicate the observational density and entropy profiles (scaled) for CC (blue) and NCC (red) clusters (McDonald et al. 2013). At sufficiently low injection rates, all models do little (as expected). In 'Momentum', 'Thermal', and 'CR' injection, we see that when the injection is nuclear ('BH') and large enough, explosive behaviour results (expelling nearly all gas within $\sim$ 30–100 kpc, and leaving what remains very hot), in stark contrast with observations. Quasi-stable intermediate cases do exist, for turbulent stirring and CR injection in particular. Among the turbulent runs with suppressed SF, most preserve the initial CC features (though they do suppress the density, heat up, and flatten the entropy profile in the core), though 'core-5/6' resemble NCC clusters (but do not 'explode'); uniform turbulent driving suppresses densities even at $\gg$ 100 kpc as well. The 'Th-core-44' run, which has non-explosively suppressed SFR, broadly resembles NCC clusters, but its negative temperature gradient is in tension with observations. The 'CR-BH-43' run, which also has a non-explosively suppressed SFR, falls between CC and NCC.

density beyond 100 kpc is also suppressed by almost a factor of 10, i.e. the halo begins to expel/lose a significant amount of gas. Among the runs with significantly suppressed SFRs, the density and entropy profiles of 'Turb-core-5' and 'Turb-core-6' end up resembling those observed in NCC clusters (compare Sanderson, Ponman & O'Sullivan 2006; Sanderson, O'Sullivan & Ponman 2009; Hudson et al. 2010; McDonald et al. 2013), while 'Turb-core-4' lives between CC and NCC. The other turbulent stirring runs with moderately suppressed SFRs preserve the CC features, although their densities in the core regions are slightly higher than observational values.

The effects of thermal heating on the halo properties strongly correlate with the kernel size of the injection. When concentrated in the black hole neighbourhood, $\sim 10^{43}\,{\rm erg\,s^{-1}}$ is sufficient to significantly suppress the density within 5 kpc and heat up the gas up to 10^8 K. Thermal injection rates in the BH neighbourhood

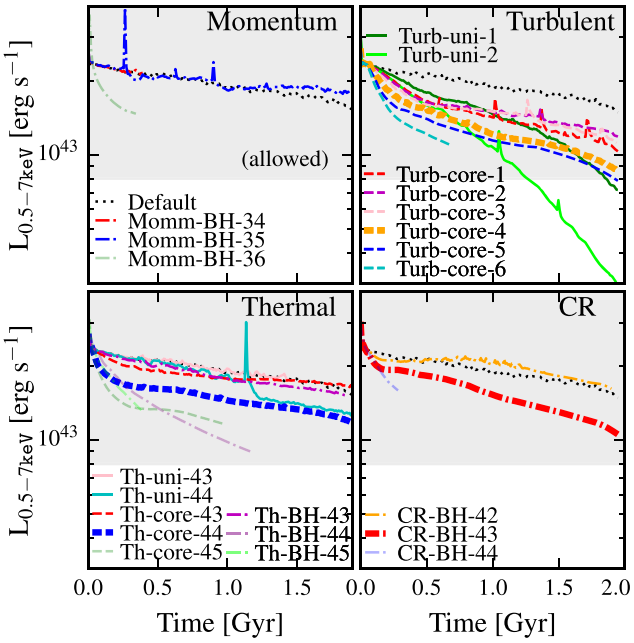

Figure 4. X-ray cooling luminosity L_X, integrated from 0.5 to 7 keV, in the runs in Fig. 3. The shaded regions indicate the observed X-ray luminosities in Reiprich & Böhringer (2002) and Stanek et al. (2006) for haloes with $m_{\rm halo} \sim 0.7$–$1.5 \times 10^{14}\,{\rm M}_\odot$. Runs that explosively eject core gas (e.g. 'Momm-BH-36', 'Th-BH-44,45', and 'CR-BH-44') strongly suppress L_X. Uniform turbulent stirring ('Turb-uni') also suppresses L_X strongly by ejecting gas (at larger radii). But other runs with suppressed SF ('Turb-core-X', 'Th-Core-44', 'CR-BH-43') have only factor ∼1.5–3 lower L_X. This is because a large portion of the total X-ray luminosity is from larger radii, although the surface brightness decays as a function of radius.

$\gtrsim 10^{44}\,{\rm erg\,s^{-1}}$ blow out everything within 10 kpc, heat gas to $\gtrsim 10^{10}$ K, and produce a negative temperature slope out to >100 kpc. If the injection is smoothed over a Gaussian kernel of 30 kpc, then the core density is not suppressed until the total energy input reaches $\gtrsim 10^{44}\,{\rm erg\,s^{-1}}$ (when the energy input is comparable to the cooling). Although milder, a negative temperature gradient extending from 10 to 100 kpc is still hard to avoid in that case.

CR injection can significantly suppress the core density with $\dot{E} \gtrsim 10^{43}\,{\rm erg\,s^{-1}}$, and produces an extended region with significant hot gas. If the input exceeds $10^{44}\,{\rm erg\,s^{-1}}$, the injection becomes explosive on large scales (similar to high-$\dot{E}$ BH-kernel thermal injection). Except for the explosive one, runs with CR injection have density and entropy profiles resembling those observed in CC clusters. The gas of the most successful CR injection run ('CR-BH-43') within ∼7 kpc is dominated by the hot gas from stellar mass-loss (and/or gas heated by CR) and is less constrained by the observations.

3.2.2 *X-ray luminosities*

The resultant X-ray luminosity of the gas halo is an important constraint for an AGN feedback model (e.g. Choi et al. (e.g. McCarthy et al. 2010; Choi et al. 2015). Fig. 4 shows the predicted X-ray cooling luminosity, integrated over all gas in the halo, from 0.5 to 7 keV. The luminosity is calculated using the same methods in Schure et al. (2009) and Ressler, Quataert & Stone (2018), in which the cooling curve is calculated for the photospheric solar abundances (Lodders 2003), using the spectral analysis code SPEX (Kaastra, Mewe & Nieuwenhuijzen 1996) and scaled according to the local hydrogen, helium, and metal mass fractions. The shaded regions indicate the observed X-ray luminosities in Reiprich & Böhringer (2002) and Stanek et al. (2006) for haloes with $m_{\rm halo} \sim 0.7$–$1.5 \times 10^{14}\,{\rm M}_\odot$. Runs that quench by violently ejecting gas strongly suppress their X-ray luminosities, as does the 'uniform' turbulent stirring run (owing to its suppression of gas densities everywhere in the halo). But interestingly, other runs with suppressed SF/cooling flows maintain X-ray luminosities just a factor ∼1.5–3 lower, well within the observed range (Reiprich & Böhringer 2002; Balogh et al. 2006; Stanek et al. 2006; Kim & Fabbiano 2013; Anderson et al. 2015). This is because a large portion of the total X-ray luminosity is from larger radii, although the surface brightness decays as a function of radius.

3.2.3 *Turbulent Mach number*

Fig. 5 shows the rms 1D turbulent velocity, defined as $v_{\rm turb}/\sqrt{3}$, and the 1D Mach number for gas hotter than 10^7 K as a function of radius, averaged over the last 100 Myr of the runs. Radial momentum injection does not alter turbulence much, as it primarily drives coherent motion; likewise for thermal injection when it is weak or spread over large radii. In the 'explosive' regime of momentum/thermal/CR input, all drive strong outflows at up to $\sim 1000\,{\rm km\,s^{-1}}$, though the higher shocked-gas temperatures mean this corresponds to Mach ∼ 0.4. At intermediate CR injection rates, appreciable but modest bulk motions are driven at $\gtrsim 10$ kpc.

By construction, turbulent stirring boosts turbulent velocities where injected. The maximum turbulent velocities reach ∼ 200–250 km s^{-1} (Mach $\lesssim 0.5$) in the 'Turb-uni' and 'Turb-core-1-4' runs, broadly consistent with observations of the Perseus cluster (Hitomi Collaboration 2016, 2018), but towards the higher end of the allowed range, while the value for 'Turb-core-5' ($\lesssim$400 km s^{-1}) is slightly higher than the observations, and 'Turb-core-6' has Mach number > 1.

3.3 Cooling time and gas stability

Fig. 6 shows the cooling time ($\tau_c \equiv E_{\rm thermal}/\dot{E}_{\rm cooling}$) versus radius for gas hotter than 10^5 K. Momentum injection does not affect the cooling time strongly. Even in our highest momentum flux run ($\sim 10^{36}\,{\rm g\,cm\,s^{-2}}$), where everything within 70 kpc is blown away, the cooling time at even larger radii still remains very similar to the 'Default' run.

On the other hand, turbulent stirring, which effectively suppresses the gas density, can also suppress the cooling rate (through turbulent mixing and pressure support). The regions with boosted cooling time roughly coincide with the regions with strong stirring. When the stirring injects $\sim \times 10^{41}\,{\rm erg\,s^{-1}}$ ('Turb-core-4') within 100 kpc, the average cooling time of gas with $T > 10^5$ K beyond 10 kpc is boosted to $\gtrsim 10$ Gyr. In 'Turb-core-5 and 6', almost all the gas becomes stably non-cooling, consistent with their resulting NCC halo properties.

Thermal heating can significantly boost the cooling time as long as the kernel of injection is small enough (only the 'Th-BH' runs and 'Th-core' runs). The increase of the cooling time basically follows the increase in temperature discussed in Section 3.2.

CR injection boosts the cooling time within $r \sim 10$ kpc, owing to lower densities and higher temperatures inside these radii, when the injection rate is $\gtrsim 10^{43}\,{\rm erg\,s^{-1}}$.

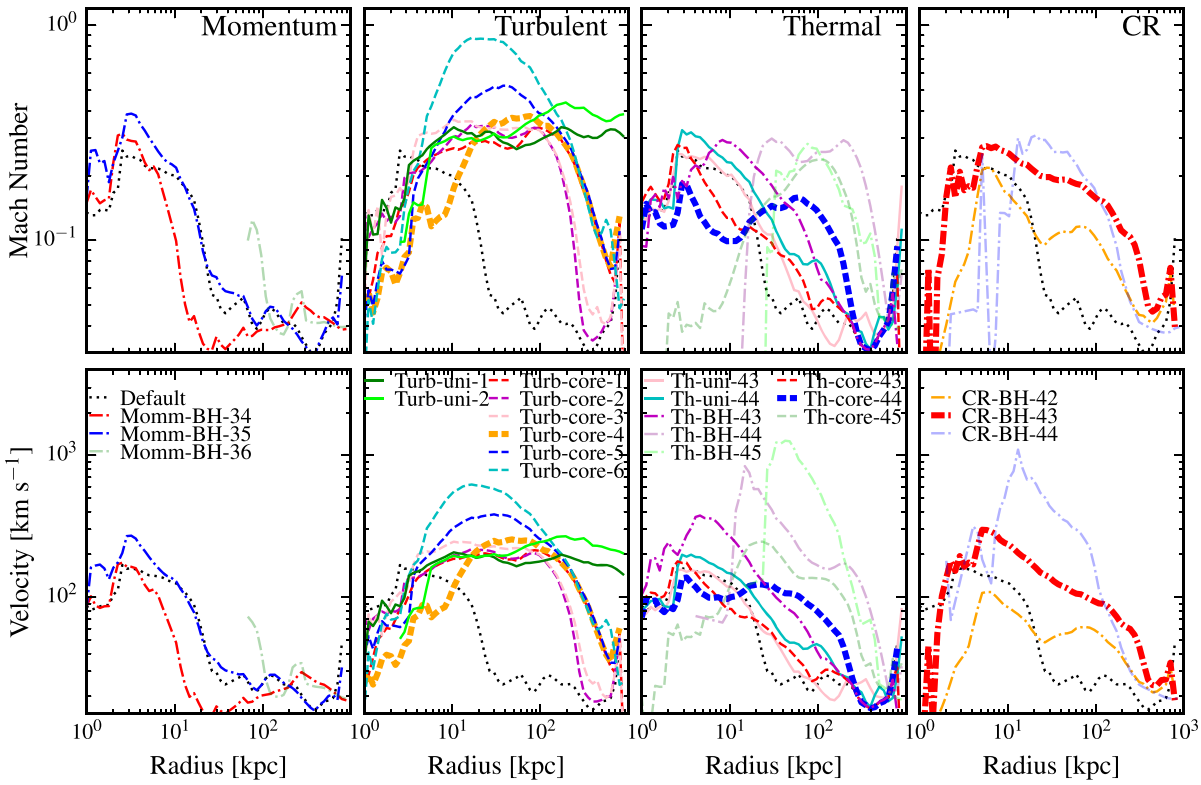

Figure 5. *Top:* 1D rms Mach number ($v_{\rm turb}/\sqrt{3}v_{\rm thermal}$, in gas with $T > 10^7$ K, averaged over the last 100 Myr of the runs) as a function of radius for the runs in Fig. 3. *Bottom:* 1D rms velocity dispersion $v_{\rm turb}/\sqrt{3}$. The 'Default' run has turbulence driven by a combination of thermal instability and stellar feedback. Weaker momentum/thermal/CR input does not alter this much; when those inputs become 'explosive' (see Fig. 3), strong shocks appear as jumps in $v_{\rm turb}$ up to $\gtrsim 1000\,{\rm km\,s^{-1}}$. Modest CR injection or distributed thermal injection contribute $\sim 200\,{\rm km\,s^{-1}}$ bulk motions at $r \gtrsim 10$ kpc. Turbulent stirring runs (by construction) produce Mach ~0.2–0.4 turbulence over the radii of the chosen kernel, although the strongest runs (e.g. 'Turb-core-6') exceed Mach $\gtrsim 1$.

The ratio of cooling time to dynamical time ($\tau_{\rm c}/\tau_{\rm d}$, with $\tau_{\rm d} \equiv (r^3/GM_{\rm enc})^{1/2}$) is also plotted as an indication of gas stability. The runs suffering from the most severe cooling flows in our suite ('Default', 'Momm-BH-34,35', 'Th-uni-43,44', 'Th-core-43', 'Th-BH-43', and 'CR-BH-42') have an extended region within 100 kpc at $\tau_{\rm c}/\tau_{\rm d} \lesssim 20$. In the runs with SFRs suppressed to $\lesssim 1\,{\rm M_\odot\,yr^{-1}}$ ('Turb-core-4', 'Th-core-44', and 'CR-BH-43'), most of the gas within this radius has $\tau_{\rm c}/\tau_{\rm d} > 10$. In the runs which end up resembling NCC clusters ('Turb-core-5 and 6'), with $\tau_{\rm c}/\tau_{\rm d} \gtrsim 100$ uniformly. Consistent with previous studies (e.g. Sharma et al. 2012; Gaspari, Brighenti & Temi 2015; Voit et al. 2017), we find that our simulations that avoid the cooling catastrophe and also produce 'realistic' CC profiles have $\tau_{\rm c}/\tau_{\rm d} \gtrsim 10$.

3.4 Energy input versus cooling

In Fig. 7, we compare cooling rates, energy input rates, and net energy gain/loss of each run, integrated within a radius r. Here 'energy input' sums stellar feedback (adding SNe and stellar mass-loss kinetic luminosities) plus the input from our analytic injection models. We also show where gas (above 10^5 K) has cooling times exceeding the Hubble time.

Direct thermal heating, as expected, suppresses cooling in the core region only if the injected heating rate is larger than cooling: this is why uniform or large-kernel heating is inefficient (energy is 'wasted' at large radii). When highly concentrated, this tends to result in explosive behaviour, which reduces the cooling rate further out not by direct heating but by ejecting the halo baryons. The only thermal heating run with heating roughly matched to cooling over the extended cooling region is the (intentionally fine tuned) 'Th-core-44' run.

Akin to the thermal runs, 'CR-BH-42' does little, 'CR-BH-44' is explosive, while 'CR-BH-43' is able to maintain quasi-stable equilibrium. A key difference is (as we show below) this comes primarily from pressure support, where the CR pressure profile (if diffusion is fast and the injection rate is constant, and losses are negligible) is essentially a steady-state $p_{\rm cr} \sim \dot{E}_{\rm cr}/12\pi\kappa\, r$. This makes the predictions less sensitive to small variations in the cooling rates or gas densities.

Turbulent stirring can suppress cooling rates significantly without becoming 'explosive' and with significantly lower energetic 'cost'. We discuss the mechanisms for this in Section 5.3.

3.5 The rejuvenation of non-cool-core clusters and role of feedback from old stellar populations

Given that 'Turb-core-5' and 'Turb-core-6' evolve from CC to NCC in a relatively 'gentle' manner, a natural question to ask is whether the halo will become CC again if the turbulent stirring is turned off. It turns out that rejuvenation does not necessarily occur, at least in these idealized simulations (remember, our simulations are non-cosmological, so do not include *new* gas accreting into the halo).

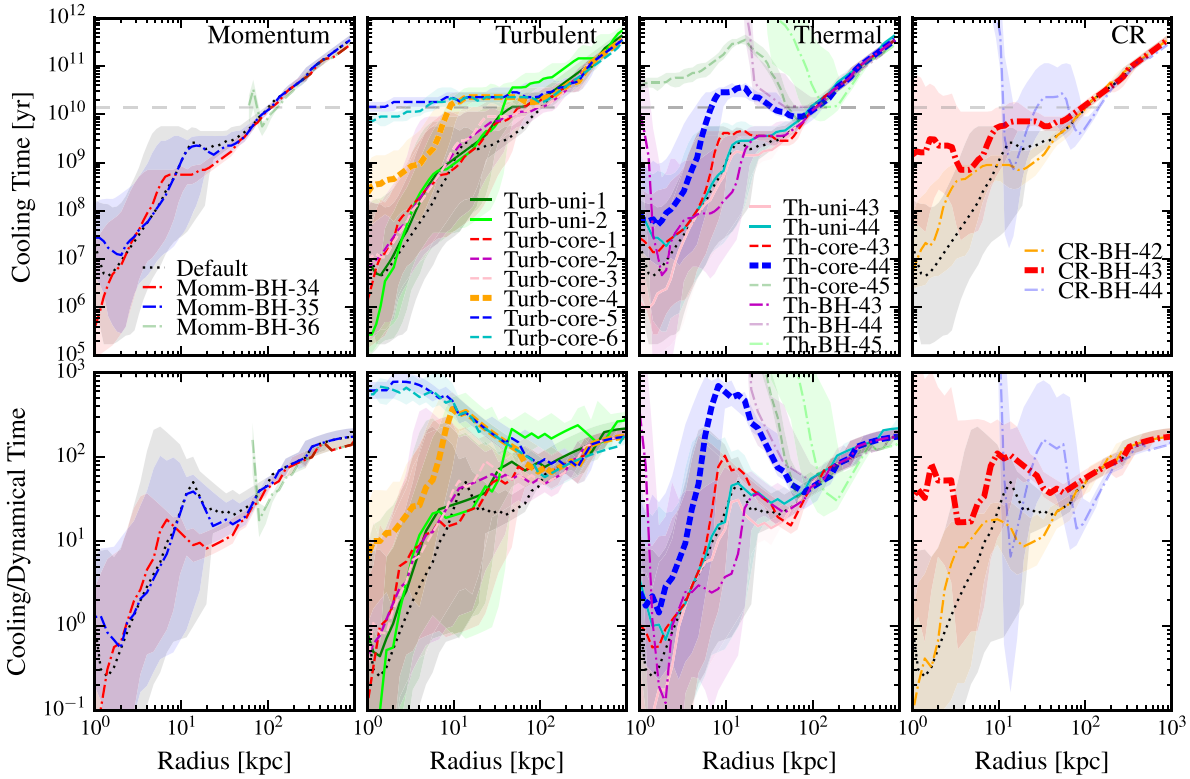

Figure 6. *Top:* Gas cooling time ($\tau_c \equiv E_{\rm thermal}/\dot{E}_{\rm cooling}$) versus radius (averaged in the last 100 Myr of the runs in Fig. 3). Grey dashed line labels the Hubble time. *Bottom:* Cooling time over dynamical time ($\tau_{\rm d} \equiv (r^3/GM_{\rm enc})^{1/2}$). With weak injection these are not strongly modified. In explosive cases the cooling time interior to the explosive shock is enormous (the 'cutoff' to zero reflects cases where there is no gas in the relevant temperature range inside some radius). CRs and turbulent stirring suppresses cooling primarily by suppressing core gas densities; regions with boosted $\tau_{\rm c}$ correspond to regions with strong stirring. If the stirring exceeds $\gtrsim 2\times10^{41}$ erg s^{-1} within 100 kpc ('Turb-core-5 and 6'), or CR injection exceeds $\gg 10^{43}$ erg s^{-1}, the gas has $\tau_{\rm c} > t_{\rm Hubble}$ and $\tau_{\rm c}/\tau_{\rm d} \gtrsim 100$ – this is an excellent predictor of when the system will resemble an NCC cluster.

We test this by restarting the 'Turb-core-5' run from the 1.4 Gyr point and the 2.0 Gyr point, removing our injection (keeping e.g. stellar feedback and all other physics, however). As shown in Fig. 8, the '1.4 Gyr' run rejuvenates (core baryonic mass slowly grows and star formation reoccurs) while the '2.0 Gyr' one does not. The reason is that once the density is lowered to a (very low) point where the residual steady-state energy input from Type 1a SNe and AGB winds surpasses cooling, the halo remains quenched for a Hubble time. Fig. 9 shows the same comparison of Fig. 7 for 'Turb-core-5' at 1.4 and 2.0 Gyr, but includes only the stellar feedback contribution in 'energy input'. It is clear that at 1.4 Gyr there is still an extended region ($r \lesssim 30$ kpc) with sufficiently dense gas that stellar feedback from old stars alone (SNe Ia and AGB mass-loss) can only marginally balance cooling, while by 2.0 Gyr, the density has been depleted to the point where the old-star stellar feedback (which is basically identical) now totally surpasses cooling.

If we restart this 2.0 Gyr run without stellar feedback from old stars (disabling Ia's and AGB mass-loss), then it does rapidly resume SF and 'rejuvenate'. We have confirmed it is the Ia population that dominates the energy injection and results here. But in either case, it appears that stellar feedback can aid in *maintaining* quenched systems, but only once they are well into the NCC stage with especially depleted central gas densities.

We focused on this case because it was only marginally an NCC cluster. In every simulations that produces 'explosive' quenching, the central gas densities are extremely low (much lower than our 2.0 Gyr run here) and so, unsurprisingly, rejuvenation never occurs.

4 RESULTS AS A FUNCTION OF HALO MASS

We now explore models in lower mass haloes **m12** and **m13** at $M_{\rm halo} \sim 10^{12}$ and 10^{13} M$_\odot$, respectively (see Table 1). We focus our attention on models motivated by those that at least seem plausibly 'successful' (able to have some effect, but also not obviously in gross violation of observational constraints) – this includes variations of the turbulent stirring 'core-kernel' runs, CR injection with appropriate energetics, and thermal heating with an appropriate-scaled spatial kernel and energy scale. The survey at high resolution is listed in Table 3, though we have run additional low-resolution tests of broader parameter space to confirm our intuition from the **m14** survey continues to hold.

For the thermal and CR injection cases, we scale the input energy from the m14 'Th-core-44' and 'CR-BH-43' runs according to the total cooling rate of the halo. For turbulent stirring, we scale the characteristic wavelength of the stirring in Fourier space (λ) and kernel size ($r_{\rm k}$) from the m14 'Turb-core-4' run according to the viral radius, and the amplitude of particle acceleration according to the circular velocity at the kernel size: $ar_{\rm k} \sim v_{\rm c}^2 \sim GM_{\rm enc}/r$. For the **m12** case, the above scaling makes the kernel very narrow and confines the stirring or energy injection to the disc, so we also included **m12** runs with a wider kernel.

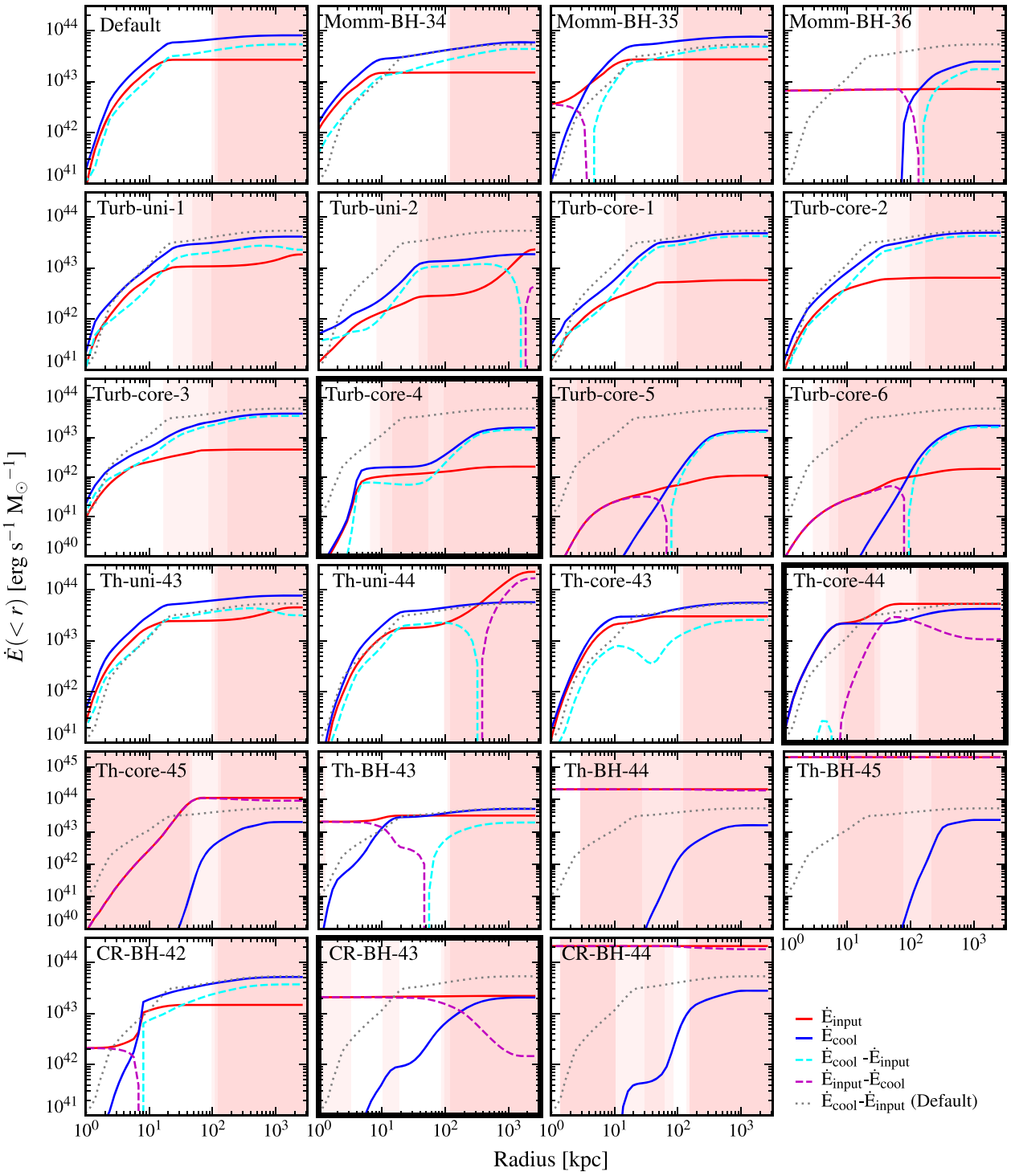

Figure 7. Cumulative (integrated inside $<r$) cooling rate ($\dot{E}_{\rm cool}$), total feedback energy input rate ($\dot{E}_{\rm input}$), and difference (net loss/gain), in the runs from Fig. 3 averaged over their last 100 Myr. Progressively darker red shading indicates regions where 20/50/80 per cent of gas (above $T > 10^5$ K) has cooling times greater than the Hubble time. Weak input produces little change; 'explosive' runs quench by dramatically lowering heating and core gas densities. Several of the turbulent runs suppress cooling significantly, ensuring $\tau_{\rm c} > t_{\rm Hubble}$, without 'exploding'. This is also seen in intermediate CR runs ('CR-BH-43'), more or less independent of the injection kernel. In thermal runs this requires an injection kernel and energy fairly carefully matched to the cooling radius/energy.

The resulting SFRs are plotted in Fig. 10. Turbulent stirring and CR injection quench all haloes, while thermal heating is less efficient in **m12**. Fig. 11 shows the density, temperature, and entropy profiles, while Fig. 12 compares the energy injection to cooling luminosities as a function of radius.

Thermal heating has similar effects in **m13** and **m14**: the galaxies are quenched but inevitably have a mild negative temperature gradient. The energy input of both matches the cooling in an extended region (by construction). However, in the **m12** case, the cooling rate is actually *boosted* by the additional thermal heating

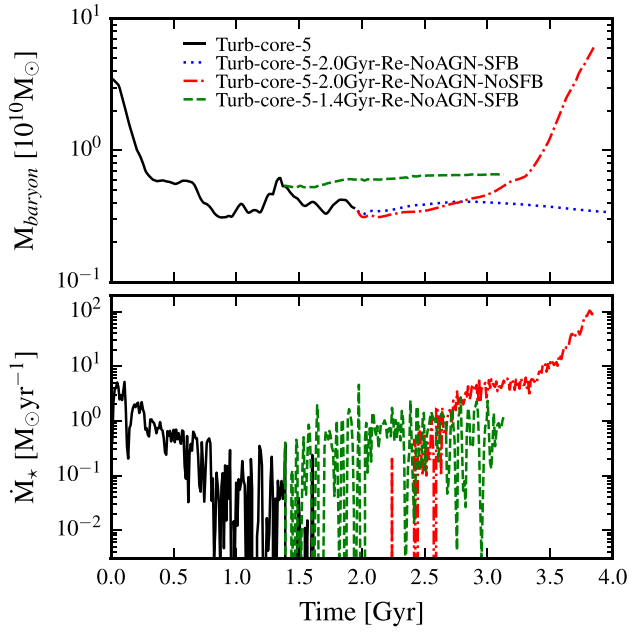

Figure 8. Testing 'rejuvenation'. We restart run 'Turb-core-5' which transitions from CC to NCC cluster, at either 1.4 Gyr ('Re-1.4Gyr-SFB') or 2.0 Gyr ('Re-2.0Gyr-SFB'), keeping all physics identical but turning off the turbulent 'stirring' at the time of restart. We compare the baryonic mass within $r < 30$ kpc (excluding pre-existing stars; *top*) and SFR (*bottom*) as Fig. 2. The earlier restart 'rejuvenates' after additional energy injection is disabled, and $\dot{M}_{\rm ast}$ slowly regrows over $\sim$ Gyr time-scales. The later restart fails to rejuvenate, as in the intervening time, continued driving has lowered the core gas density to the point where stellar feedback from old stars (Ia and AGB) can keep it hot. We confirm the latter by rerunning the 2.0 Gyr restart without this feedback ('Turb-core-5-2.0Gyr-Re-NoAGN-NoSFB'), which now rejuvenates.

(because the virial temperature is low and amount of gas at $\sim 10^4$–10^5 K is large, this pushes gas higher on the cooling curve), so the effect is much weaker. In lower resolution tests, this is not remedied by increasing the thermal energy injection rate: because of the more violent thermal instability, all our thermal-heating runs in **m12** either produce no effect or violent explosion of the entire halo.

Scaling the turbulent 'stirring scale' with the viral radius (and strength with the circular velocity) stably quenches **m13**. In **m12** these naive scalings lead to stirring confined to $\sim$16 kpc, which effectively stirs the galactic disc and ballistically 'launches' the whole disc into fountains, which produce a violently bursty star formation history. Obviously this is not realistic: increasing the 'stirring kernel' size to $\sim$40 kpc ('m12-turb-core-wide') produces a much smoother low SFR and stable CC structure with slightly lower core densities and cooling rates.

CR injection successfully and 'smoothly' quenches **m12** and **m13**. In **m13**, the overall cooling rate is eventually also suppressed significantly as the gas within the central few kpc is ejected.

5 DISCUSSION: HOW DO DIFFERENT PHYSICS QUENCH (OR NOT)?

The only injection models that result in a semistable quenched galaxy are thermal heating with a Gaussian kernel chosen in the correct energy and size range, turbulent stirring confined to radii below the halo scale radius, and CR heating in the correct energy range. We summarize the key properties of these 'more successful' runs in Fig. 13. We then briefly discuss how each surveyed model in Section 2.2 operates.

5.1 Radial momentum injection

The actual kinetic energy (even if it all thermalized) of the momentum-injection runs is less than the cooling luminosity.[11] At low injection rates this just stirs small-scale turbulence/fountains (e.g. in 'Momm-BH-35', 40 per cent of the momentum and 20 per cent of the energy input is used to decelerate in-falling gas, and $\sim$1/2 of the gas acted directly upon by the nuclear stirring still forms stars). At higher injection rates it acts by dynamically altering halo structure, ejecting material from the core. Without truly enormous energy input this is eventually decelerated in the outer halo, but in e.g. 'Momm-BH-36' almost all the gas within $\lesssim$ 70 kpc is ejected.

Previous studies have similarly noted that pure isotropic kinetic input tends to fall into burst-quench cycles where either it fails to alter the cooling flow or explosively ejects all the gas in the cooling radius (Ciotti et al. 2009; Shin, Ostriker & Ciotti 2010). One alternative is to inject energy in a completely different form (discussed below). A second is to inject momentum at larger radii (distributing it away from the centre), in a spatially localized-but-time-dependent manner (i.e. not in a simple radially outward-moving shell, which simply repeats this problem on larger scales; see Gaspari et al. 2011; Li & Bryan 2014; Yang & Reynolds 2016; Bourne & Sijacki 2017 for kinetic jets) – this is much closer in practice to our 'turbulent stirring' runs below. A third alternative is to invoke a mix of isotropic kinetic feedback and thermal feedback (as in Ciotti, Ostriker & Proga 2010; Dubois et al. 2013; Weinberger et al. 2017b; Pillepich et al. 2018), a possibility we discuss below.

5.2 Thermal heating

In pure thermal heating models, nothing changes unless the overall heating rate is larger than the cooling rate. However, unless these are carefully balanced, this tends to produce a negative temperature gradient in direct contradiction with observed systems (Brighenti & Mathews 2002; Mathews, Faltenbacher & Brighenti 2006), and can drive explosive behaviour that removes most of the gas in the halo. As a result, we must tune the energy input to match the cooling rate. We must also tune the injection radius to match the cooling radius, or else the energy is either 'wasted' on gas at large radii (not cooling efficiently) or it excessively heats gas in the centre driving Sedov–Taylor blastwaves that heat gas to very high temperature, eject gas in the central halo, produce negative temperature gradients, and strong shocks.

This is also consistent with previous studies that have repeatedly found nuclear energy injection alone tends to either fail to quench or violently eject far too much gas from haloes (Genel et al. 2014). The alternatives are typically to invoke either (1) fine tuning or (2) some mix of other feedback mechanisms.

5.3 Turbulent stirring

In almost all of our turbulent stirring runs that produce suppressed cooling flows, the turbulent energy injection rate is much lower than

[11] The total kinetic energy input roughly matches the total cooling rate at the begin of the most explosive 'Momm-BH-36' case, but soon become subdominant.

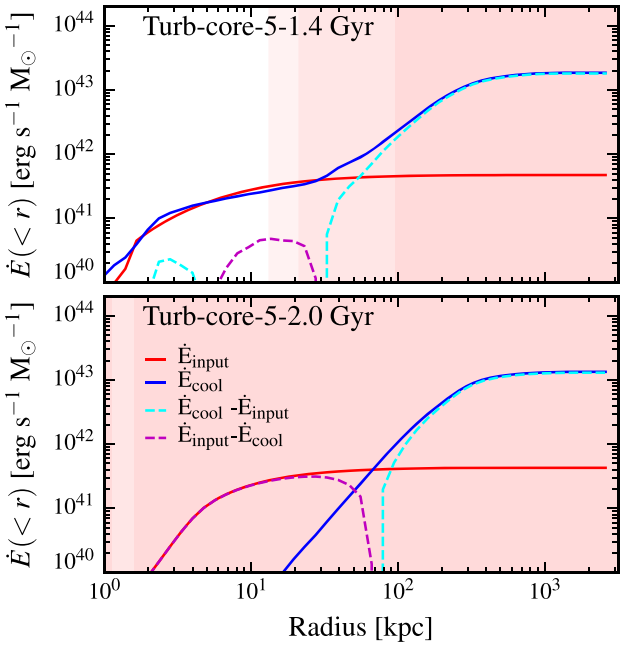

Figure 9. Cumulative cooling rate versus energy input (as Fig. 7, but including *only* stellar feedback from old stellar populations in the $\dot{E}_{input}$ budget), of the 'restarted' runs in Fig. 8. In the earlier restart, the higher core densities allow cooling and rejuvenation after the turbulent injection is deactivated. In the later restart, the stellar injection is identical (it comes from the same old stars) but the gas density has been further lowered by the injection to the point where old stars can maintain quenching for a Hubble time.

the total cooling rate (especially pre-turbulence) – in other words, thermalized turbulent energy 'heating' gas is not the dominant channel. We have also directly confirmed this by measuring the turbulent damping rate and comparing it to cooling. More important, turbulence mixes gas to larger radius, which (i) lowers the central density, (ii) lowers the density of 'up-welling' parcels (lowering their cooling rates), and (iii) mixes them with hot gas (providing a form of 'bulk conduction'). Together this lowers the effective cooling rate by an order of magnitude in the region with cooling times shorter than the Hubble time.

In turbulent mixing, cool gas moves out, while hot gas moves in. These processes work together to destroy and re-mix the cool gas at larger radii. This brings the cool gas in contact with a much larger thermal energy reservoir. For instance, Fig. 14 tracks the evolution of gas that is initially cold and dense in 'Turb-core-4'. After ~ 0.5 Gyr, less than half of this gas remains cold and dense or forms stars: most is instead shifted to a larger radii and mixed with hot gas. This is consistent with the picture proposed in Cho et al. (2003), Voigt & Fabian (2004) and other studies (e.g. Kim & Narayan 2003; Vernaleo & Reynolds 2006; Parrish et al. 2010; Ruszkowski & Oh 2010, 2011; Banerjee & Sharma 2014).

However, we emphasize that, as others have found (e.g. Voigt & Fabian 2004; Parrish, Quataert & Sharma 2009; Smith et al. 2013), it is not sufficient to consider turbulence as an 'effective conductivity' (Paper I). This is because the kinetic bulk motion of cold gas elements is important in a thermally unstable stratified medium. Moreover, unlike the effective conductivity treatment, turbulent stirring does not preserve an exactly steady-state power-law density profile. The central densities can be suppressed, including to values within the range observed for real clusters (Figs 3, 4, 6, and C1), which non-linearly decreases the cooling rate.

Also, unlike the 'effective conductivity' treatment, which tends to flatten abundance profiles (see e.g. Rebusco et al. 2005, 2006; Graham et al. 2006), turbulent stirring by itself does not appear to change the abundance profiles significantly. Instead, as discussed in Appendix B, the resulting density profiles, SFRs, and fraction of metals that are recycled can have larger effects on the abundance profiles.

Table 3. Physics variations (run at highest resolution) in our survey of lower mass (**m12** and **m13**) haloes.

Model	λ (kpc)	$\dot{E}_{tot}$ (erg s^{-1})	$\dot{P}_{tot}$ (g cm s^{-2})	Kernel (r in kpc)
m12-Turb-core	25	5.5–8.4 e39	4.0–4.2 e32	$a_{r<20} \sim 3\exp(-(r/15.8)^2)$ $a_{r>20} \sim \exp(-r/39.6)$
m12-Turb-core-wide	25	4.7–9.3 e39	4.6–4.9 e32	$a_{r<50} \sim 3\exp(-(r/39.6)^2)$ $a_{r>50} \sim \exp(-r/100)$
m13-Turb-core	57	1.3–1.6 e40	10–8.6 e32	$a_{r<50} \sim 3\exp(-(r/40)^2)$ $a_{r>50} \sim \exp(-r/100)$
m14-Turb-core-4	120	5.7–5.9 e41	3.1–2.0 e34	$a_{r<100} \sim 3\exp(-(r/79)^2)$ $a_{r>100} \sim \exp(-r/200)$
m12-Th-core-43	–	1.3–1.4 e43	–	$\dot{E} \propto \exp(-(r/6)^2)$
m12-Th-core-43-wide	–	1.3–2.5 e43	–	$\dot{E} \propto \exp(-(r/14)^2)$
m13-Th-core-43	–	17–8.8 e42	–	$\dot{E} \propto \exp(-(r/14)^2)$
m14-Th-core-44	–	2.0–0.5 e44	–	$\dot{E} \propto \exp(-(r/30)^2)$
m12-CR-BH-42	–	1.3e42	–	BH neighbour
m13-CR-BH-42	–	1.8e42	–	BH neighbour
m14-CR-BH-43	–	2.1e43	–	BH neighbour

Note. Partial list (including just simulations at 'production' resolution) of runs in haloes **m12** and **m13**. Style is identical to Table 2, but we add one column λ, denoting the wavelength of the turbulent driving modes. We focus only on models which were successful without being 'explosive' in the **m14** suite, and scale the energetics and kernel sizes with the cooling luminosities and virial radii, respectively. We have run additional low-resolution tests akin to the suite in Table 2 to confirm much larger/smaller injection produces similar results to what is seen there.

5.4 Cosmic ray injection

Like turbulent stirring, CR injection provides another source of non-thermal pressure support, so the gas density and the cooling rate can be suppressed without directly heating up the gas. As shown in Fig. 15, CR energy does contribute as an important pressure source, reaching at least equipartition to the thermal energy. Moreover, CR diffusion spreads out the energy input and forms a quasi-steady-state isotropic pressure gradient even if all the CR energy is injected in the vicinity of the black hole.

Fig. 15 shows that only a small fraction of the CR energy is thermalized as the CRs propagate in the 'CR-BH-43/44' models,[12] and in all cases the heating from CRs is well below total cooling rates. This is expected: the time-scale for CRs to lose energy to hadronic + Coulomb processes is $\sim 30\,\mathrm{Myr}\,(n/\mathrm{cm}^{-3})^{-1}$ while the diffusion time-scale is $\sim r^2/\kappa_{\rm cr}$, so for our parameters losses in the core are only significant if its mean density exceeds $n \gtrsim 0.1\,\mathrm{cm}^{-3}\,(r_{\rm core}/10\,\mathrm{kpc})^{-2}$.[13] However, in 'CR-BH-43/44', the temperature in the very centre ($\lesssim$ 5–10 kpc) does become large: this owes to CR pressure gradients suppressing the nuclear gas density sufficiently so that the low-density gas is heated efficiently by stellar feedback from the bulge and CR streaming heating.

On the other hand, the CR pressure gradient in Fig. 15 is able to offset gravity. If losses are negligible and diffusion dominates transport, around a point source with constant $\dot{E}_{\rm cr}$, the equilibrium pressure profile (assuming CRs are a $\gamma = 4/3$ ultrarelativistic fluid) is $P_{\rm cr} = \dot{E}_{\rm cr}/12\pi\,\kappa\,r$, which agrees well with the inner parts of our CR runs (outside the 'holes' in 'CR-BH-43/44' within the central few kpc, where stellar feedback dominates).[14] Comparing this to the gravitational force we have

$$\frac{F_{\rm CR}}{F_{\rm G}} = \frac{1}{3\rho}\frac{\partial e_{\rm CR}/\partial r}{GM_{\rm enc}/r^2} \sim 2\left(\frac{\dot{E}}{10^{43}\,\mathrm{erg\,s^{-1}}}\right)\left(\frac{\kappa}{10^{29}\,\mathrm{cm^2\,s^{-1}}}\right)^{-1}\left(\frac{r}{10\,\mathrm{kpc}}\right)^{-1} \left(\frac{v_c}{500\,\mathrm{km\,s^{-1}}}\right)^{-2}\left(\frac{n}{0.01\,\mathrm{cm^{-3}}}\right)^{-1}. \quad (1)$$

This is consistent with our result that when the CR energy input reaches $\sim 10^{43}\,\mathrm{erg\,s^{-1}}$, the CR pressure gradient starts to surpass the gravitational force in the core region and the core density and cooling rate start to be suppressed.

We emphasize that, as shown in Fig. 15, the heating from streaming loss does exceed the cooling rate in the core region in 'CR-BH-43' (the most stably quenched CR injection case), which is consistent with the previous studies (e.g. Ruszkowski et al. 2017b). However, we argue that the quenching is majorly caused by CR pressure lowering the gas density and therefore also the cooling rate instead of CR heating overcoming cooling because (i) the cooling rate of 'CR-BH-43' is much lower than the 'Default' run; (ii) the heating from stellar feedback can be at least comparable to the CR heating in the core region; and (iii) the black hole thermal heating run with exactly the same energy input does not quench the galaxy.

Given that we are not directly balancing the cooling rate by CR heating, the total CR energy in the halo does not need to be excessively high. The estimated > GeV gamma-ray luminosity of 'CR-BH-43' (from hadronic loss) is $L_\gamma \sim 10^{41}\,\mathrm{erg\,s^{-1}}$, which is lower than the observational upper bounds (Ackermann et al.

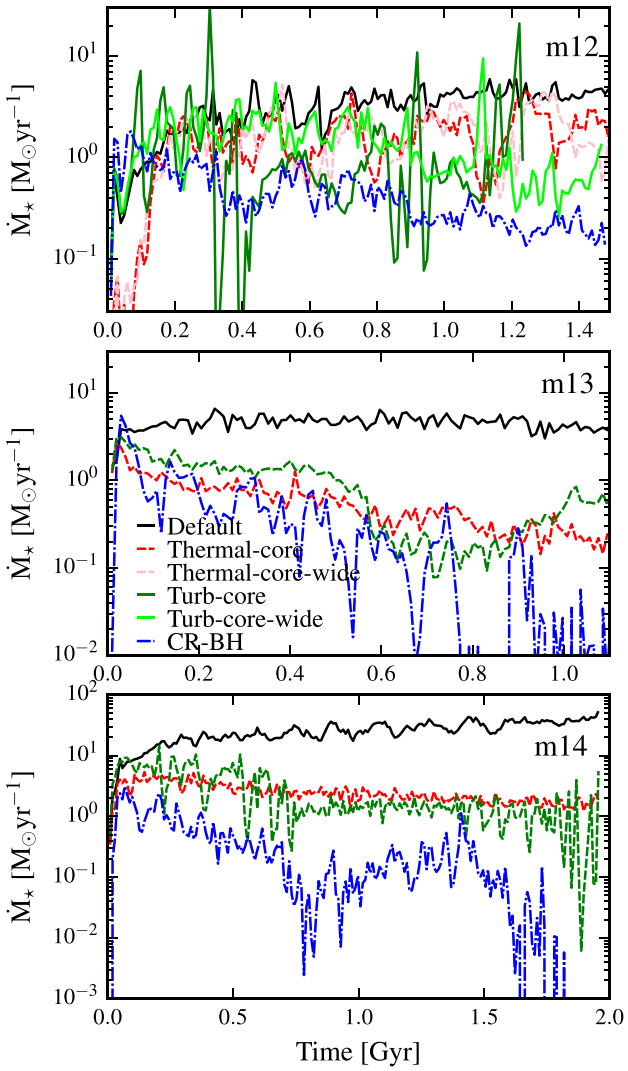

Figure 10. Galaxy SFRs (as Fig. 2) in our suite of simulations of different mass haloes (Table 3). The 'wide' runs are only for m12. In **m13** and **m14**, turbulent stirring within the halo scale radius, or CR injection with appropriate energies, can quench, as can somewhat fine-tuned thermal energy injection. **m12** is more unstable and we find no thermal-heating solutions that quench without explosive ejection of halo baryons. Also in **m12** the 'Turb-core' run confines stirring to ∼16 kpc, effectively 'churning' the galactic disc and producing the bursty star formation; this disappears with a more extended stirring ('Turb-core-wide').

[12]The total thermalized CR energy as indicated by the green and magenta lines in the second row of Fig. 15 is less than 10 per cent of the total CR input.

[13]This explains why the run 'CR-BH-42', which does not quench and maintains dense gas in the centre, does lose a non-negligible fraction ∼1/2 of its CR energy to collisional + streaming losses. For 'CR-BH-43' and 'CR-BH-44', the collisional loss is more significant initially, but it drops to a lower value after the core density is suppressed. The competition between CR energy and gas densities being larger at small r, and diffusion times longer at large r, also explains why the collisional + streaming losses have the broad radial structure seen in Fig. 15.

[14]At large radii, if the streaming is at a quasi-constant Alfv´en speed, streaming will dominate over diffusion at $r \gtrsim \kappa_{\rm cr}/v_{\rm stream} \sim 30\,\mathrm{kpc}\,(\kappa_{\rm cr}/10^{29}\,\mathrm{cm^2\,s^{-1}})\,(30\,\mathrm{km\,s^{-1}}/v_{\rm stream})$, which also defines the radius where streaming losses $\propto v_{\rm stream}\,\partial P_{\rm cr}/\partial r$ will be largest. Note in the simulations here the 'cutoffs' in the CR profiles at $r \sim$ 100–1000 kpc owe to the simulations only having finite time for CRs to propagate from the nucleus to large radii.

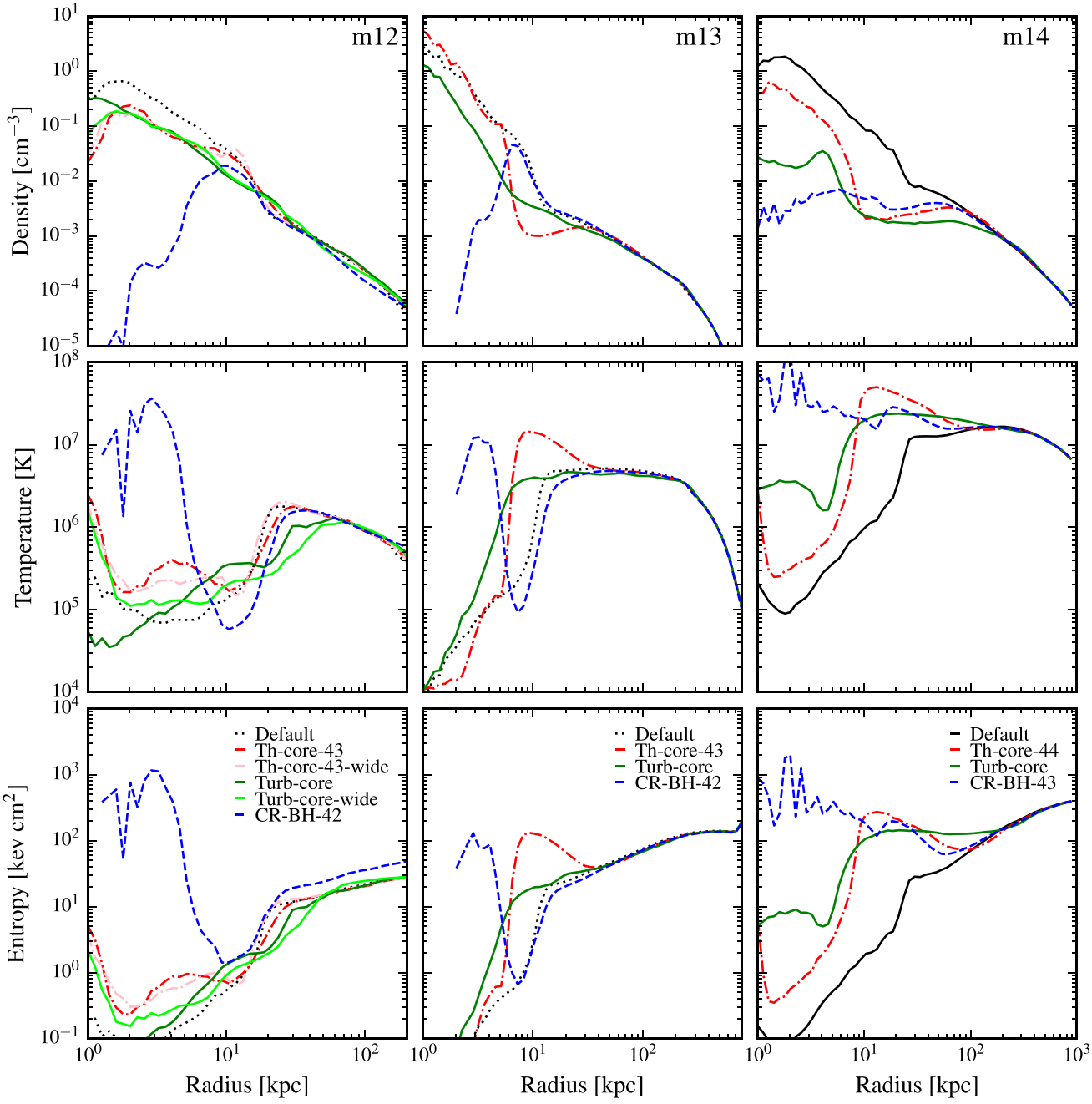

Figure 11. Density, temperature, and entropy profiles (as Fig. 3 but mass weighted) averaged over the last $\sim$ 100 Myr of each run for the runs in Fig. 10. Even fine-tuned thermal heating produces negative temperature gradients in **m13** and **m14** and has little effect in **m12**. Turbulent stirring significantly depresses the density of **m13** (and raises its temperature), resembling the stronger stirred **m14** cases, but a weaker stirring amplitude alleviates this. CR injection suppresses the core density inside the central few kpc (leading to mostly hot gas inside this radius), but leaves a positive temperature profile and intact density profile outside $r > 5$–10 kpc.

2016; Wiener & Zweibel 2019). Besides, the estimated $\sim$ GHz radio luminosity of 'CR-BH-43' from the secondary CR electrons (from CR protons), which contributes as part of the overall radio luminosity, is $L_{\rm radio} \sim 10^{39}\,{\rm erg\,s^{-1}}$, again within the observational constraint from the radio flux (e.g. Giacintucci et al. 2014; Bravi, Gitti & Brunetti 2016).[15]

[15]We assume that all the secondary CR electrons decay via synchrotron emission.

6 CONCLUSIONS

In this paper, we have attempted a systematic exploration of different qualitative physical mechanisms by which energy can be injected into massive haloes to quench galaxies and suppress cooling flows. We specifically considered models with radial momentum injection (e.g. 'wind' or 'radiation pressure' or 'isotropic kinetic' models), thermal heating (e.g. 'shocked wind' or 'isotropic sound wave' or 'photo/Compton heating' or 'blastwave' models), turbulent 'stirring' (e.g. 'convective/buoyant bubble' or 'precessing jet' or 'jet/bubble instability-driven' or 'subhalo/merger/satellite wind-

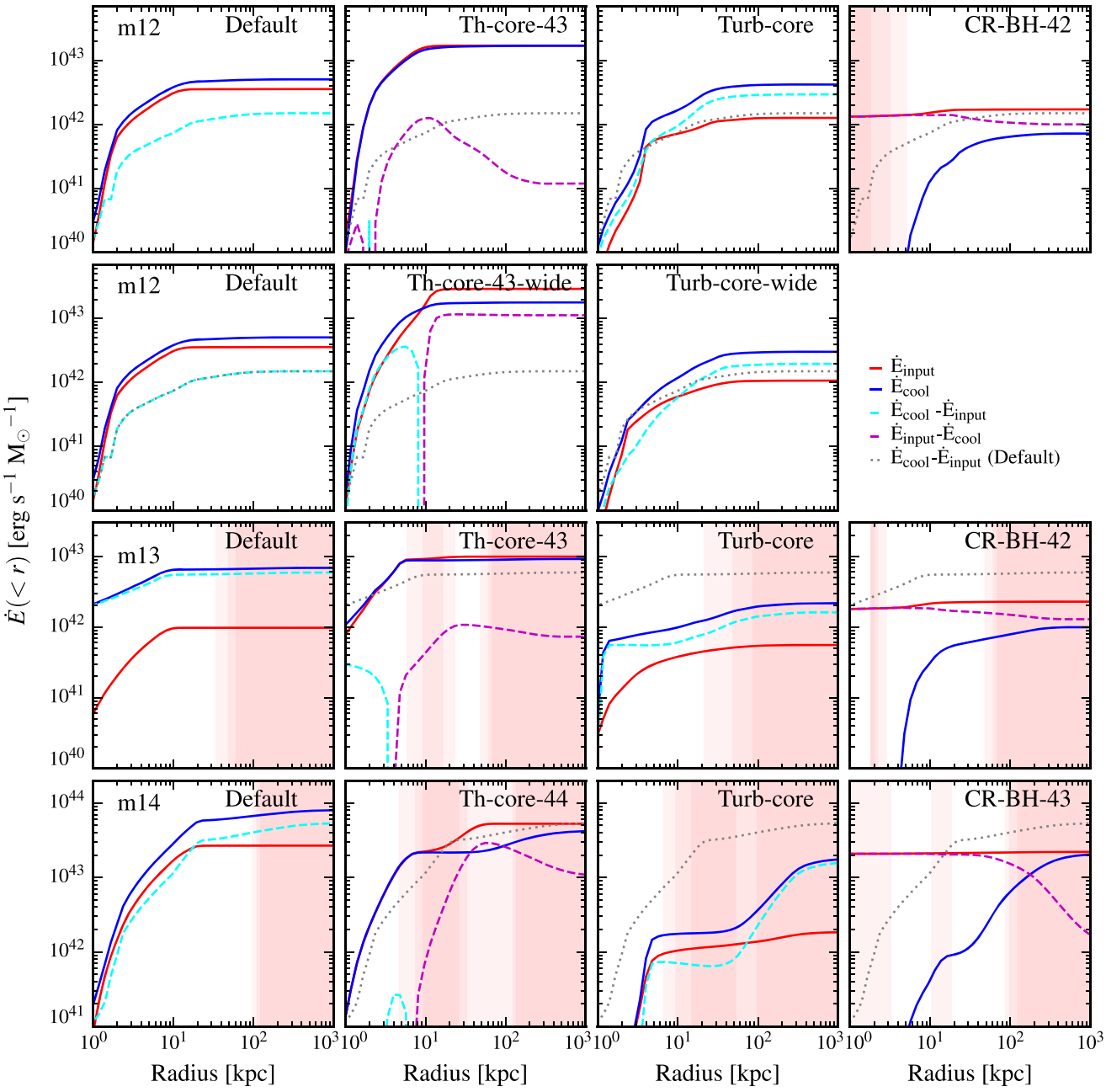

Figure 12. Cumulative energy input versus cooling (as Fig. 7) for the simulations in Fig. 10. Turbulent stirring significantly decreases cooling in **m13** (owing to lower densities in Fig. 11; this is sensitive to the injection rate), but only suppresses cooling in the core in **m12**. Thermal heating can marginally balance cooling in **m13** and **m14** which have hotter, more stable gaseous haloes, but in **m12** heating puts more dense + enriched gas at $\sim 10^4$–$10^{5.5}$ K, increasing its cooling rate. CR injection substantially suppresses cooling rates in all cases.

driven' models), and CR injection (e.g. CRs from compact or extended radio jets/lobes, shocked disc winds, or inflated bubbles). We vary the associated energetics and/or momentum fluxes, spatial coupling/driving scales, and halo mass scale from $\sim 10^{12}$ to 10^{14} M$_\odot$. These were studied in fully global but non-cosmological simulations including radiative heating and cooling, self-gravity, star formation, and stellar feedback from SNe, stellar mass-loss, and radiation, enabling a truly 'live' response of star formation and the multiphase ISM to cooling flows; we used a hierarchical super-Lagrangian refinement scheme to reach $\sim 10^4$ M$_\odot$ mass resolution, much higher than many previous global studies.

Of the cases surveyed, only turbulent stirring within a radius of order the halo scale radius, or CR injection (with appropriate energetics) were able to maintain a stable, CC, low-SFR halo for extended periods of time, across all halo masses surveyed, without obviously violating observational constraints on halo gas properties or exceeding plausible energy budgets for low-luminosity AGNs in massive galaxies.

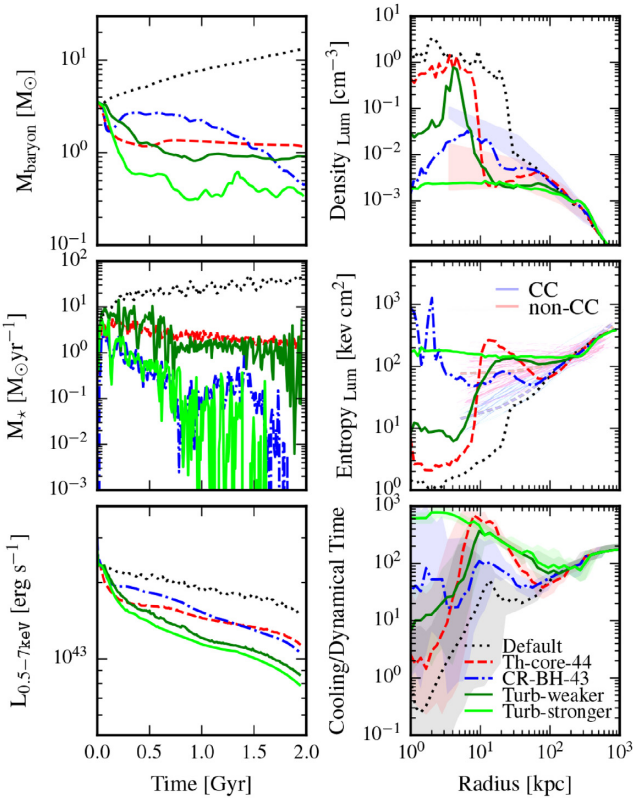

Figure 13. Executive summary: We summarize the key properties of the more successful m14 runs, 'Th-core-44', 'CR-BH-43', 'Turb-weaker' ('Turb-core-4'), and 'Turb-stronger' ('Turb-core-5'), in comparison with the stellar feedback-only run ('Default'). On the left, we show the baryonic mass within 25 kpc, SFR, and X-ray luminosity as a function of time. On the right, we show the luminosity-weighted density, entropy, and cooling time over dynamical time profiles averaged over the last 100 Myr. All these 'more successful' runs produce quenched central galaxies. The total baryonic mass in the core regions is stably reduced relative to the 'Default' run, and the runs also have enhanced central entropy. The 'Turb-core-5' run is completely converted into an NCC system, while the others fall between CC and NCC. The thermal heating run has a negative entropy gradient from a few 10s to ∼100 kpc, in tension with the observations. The cooling time over dynamical time profiles of all the runs are overall increased relative to the stellar feedback-only run, consistent with suppressed cooling flows.

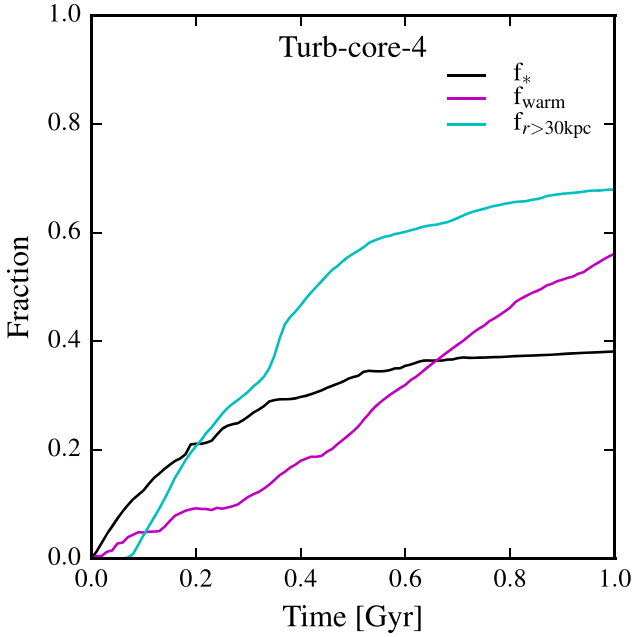

Figure 14. An example of what happens to rapidly cooling gas in the turbulent stirring runs which suppress cooling flows. Here in 'Turb-core-4' from Fig. 3, we track gas which is cold ($T < 8000$ K) and dense ($n > 1$ cm^{-3}) at an early time (20 Myr) and follow its evolution. We first follow how much of the gas forms stars – this stabilizes as ∼ 40 per cent at late times. The majority of the gas is mixed to larger radii (here > 30 kpc), and becomes warm ($T > 10^5$ K).

(i) Isotropic momentum injection with momentum flux lower than $\sim 10^{36}\,\mathrm{g\,cm\,s^{-2}}\,(M_{\rm halo}/10^{14}\,\mathrm{M_\odot})^{1/3}$ has little effect on cooling flows or star formation, while larger momentum fluxes simply generate an 'explosion' that evacuates gas from the halo core, drives strong shocks in the outer halo, generates steep negative temperature gradients out to > 100 kpc, and heats gas to enormous temperatures (all in conflict with observations).

(ii) Thermal heating, if concentrated in the halo core, similarly transitions sharply from doing nothing when the input is below cooling rates, to generating an explosive Sedov–Taylor blastwave when the input exceeds cooling rates (again, in conflict with observations). Thermal heating extended over too large a radius 'wastes' all its energy at very large radii and does little in the core. It is possible to fine tune thermal heating (by setting energy input equal to cooling rates, and the coupling scale equal to the cooling radius), but this (i) requires thermal heating rates $\gtrsim 10^{44}\,\mathrm{erg\,s^{-1}}$ in $\gtrsim 10^{14}\,\mathrm{M_\odot}$ haloes (corresponding to bright quasars if the heating efficiency is ∼ 1 per cent), (ii) still generates mild negative temperature gradients to ∼ 100 kpc, and (iii) fails in less massive haloes $\lesssim 10^{12.5}\,\mathrm{M_\odot}$ where virial temperatures are lower.

(iii) CRs can suppress cooling and SFRs by supporting non-thermal pressure gradients which are comparable to or exceed gravity in the core, with modest energetics in an order of magnitude range around $\dot{E}_{\rm cr} \sim 10^{43}\,\mathrm{erg\,s^{-1}}\,(M_{\rm halo}/10^{14}\,\mathrm{M_\odot})$. CR 'heating' (via streaming or collisional terms) is negligible as modelled here in the interesting regime. For reasonable diffusivities, the injection scale/kernel also does not matter sensitively since CRs form an equilibrium diffusion profile, unless the injection scale is very large $\gtrsim$ 30–100 kpc. The central few kpc tend to be 'hot' because they are eventually depleted of all dense gas, but the larger scale density/temperature/entropy structure of the CC halo can be stably maintained for extended periods of time, despite suppressed SFRs and actual cooling flow rates on to the galaxy.

(iv) Turbulent stirring can also suppress star formation, through a combination of suppressing the core gas density (by providing non-thermal pressure and 'lofting' parcels up the potential where they buoyantly expand), and mixing cold and dense gas into the hot halo (providing 'bulk conduction'), with even lower energetics in $\dot{E}_{\rm turb} \sim 10^{41-42}\,\mathrm{erg\,s^{-1}}\,(M_{\rm halo}/10^{14}\,\mathrm{M_\odot})$ or (equivalently) momentum flux $\dot{P}_{\rm turb} \sim 10^{34}\,\mathrm{g\,cm\,s^{-2}}\,(M_{\rm halo}/10^{14}\,\mathrm{M_\odot})$ within a radius of order the halo scale radius ($\lesssim$ 100 kpc). Towards the low end of this range, haloes maintain CC features, while towards the high end, they evolve from CC to NCC. Strong stirring at $r \gtrsim 100$ kpc tends to remove significant gas from the halo and suppresses the X-ray luminosity below observations; stirring confined only to $\lesssim$10–20 kpc acts more like galactic fountains and fails to efficiently suppress cooling. Turbulent 'heating' (via compression or shocks or viscosity) is never dominant.

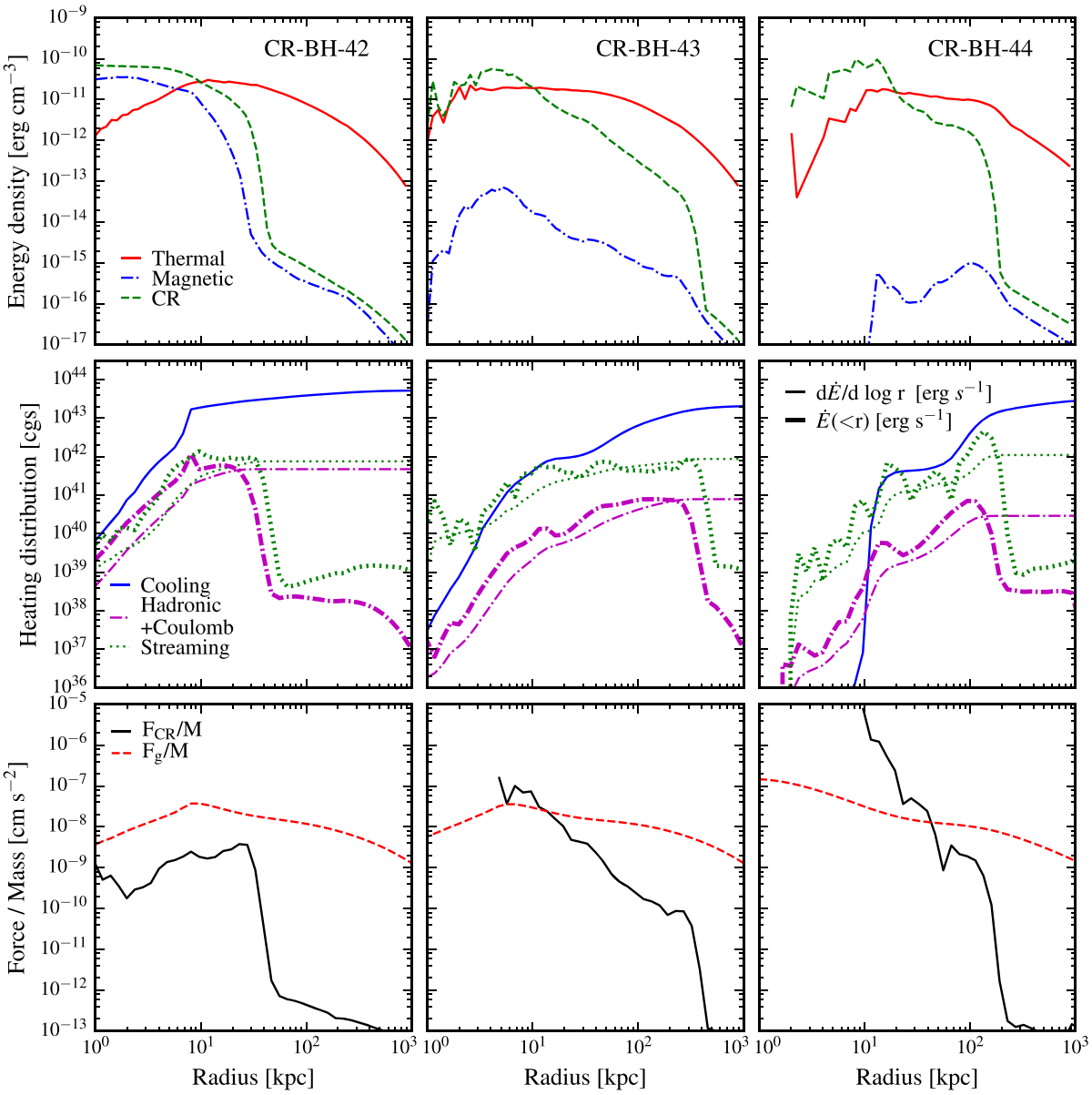

Figure 15. Comparison of energetics in our **m14** CR runs. *Top:* CR, magnetic, and thermal energy densities (averaged in spherical shells in the last 100 Myr of each run), for the three **m14** CR runs (Table 2). CR energy is non-negligible within < 30 kpc in each. *Middle:* Comparison of differential per-unit-radius ($d\dot{E}/d\log r$) and cumulative ($\dot{E}(<r)$) gas cooling rates versus CR 'heating' rates. The latter includes collisional (hadronic + Coulomb) and streaming losses, which transfer energy from CRs to thermal gas energy. CR heating is always much smaller than gas cooling except where the gas is almost completely evacuated in the central few kpc (and $\dot{E}_{\rm cool}$ is extremely small). In CR-BH-42, which retains dense central gas where CR losses are large, ~ 60 per cent of the injected CR energy is thermalized, but in the CR-BH-43/44 runs where the central few kpc are lower density, only $\sim$ 1–5 per cent of the CR energy is ever thermalized. *Bottom:* Gravitational acceleration $F_g/M \approx \partial\Phi/\partial r$ versus acceleration from the CR pressure gradient ($\rho^{-1}\,\partial P_{\rm cr}/\partial r$). CR pressure dominates and pushes material out from the central cooling core, to larger-r at larger $\dot{E}_{\rm cr}$.

(v) If injection transforms a halo into an NCC, then if the core density is suppressed to an extent that the energy input from old stellar populations (SNe Ia and AGB mass-loss) exceeds cooling rates, the halo never 'rejuvenates' even if the feedback injection shuts off.

In summary, our study supports the idea that quenching – at least of observed $z \sim 0$ massive haloes – is not dominated by single violent or 'explosive' events, but by lowering densities and suppressing cooling via mechanisms that involve relatively mild energetics and non-thermal pressure. Turbulence and CRs represent promising avenues to this, either of which has the potential to quench the models surveyed here without obviously contradicting basic observational constraints. Both operate very efficiently, with required energetics comparable to those expected in jets of low-luminosity AGNs.

We emphasize that we are not saying it is impossible to devise models of feedback using a combination of thermal and radial mechanical energy input, which produce quenching and plausible massive halo properties (in fact, we explore a couple such models here). However, consistent with most previous studies, we find that

these classes of models (i) require fine tuning, in energetics and coupling scale as a function of halo mass, (ii) generally require optimistically high energetics (at least order of magnitude larger than the CR models, and two orders of magnitude higher than the turbulent models favoured here), and (iii) may still have difficulty reproducing more subtle observational properties (e.g. distributions of temperature profile slopes).

We should also emphasize that we are not implying that AGN feedback is represented by any one of these mechanisms (especially as we model them). Real feedback is a mix of many different processes operating at once, often simultaneously on very different scales (e.g. radiation and accretion-disc winds and jets may be coupling to the gas all on different spatial scales). Our goal was simply to focus on an (intentionally) highly idealized model of each form of injection, to understand the constraints and different qualitative behaviours of different types of energy injection. This paper was a follow-up to Paper I, where we also surveyed a large number of simulations to emphasize that *something* beyond the 'default' physics of cooling, self-gravity and gravitational stability, magnetic fields, conduction, viscosity, star formation, and feedback from stars (radiative and SNe and stellar mass-loss), was required to resolve the cooling flow problem. Here, we identify plausible *classes* of physical candidates for that 'something' (e.g. enhanced turbulence and CR from AGNs). In our next study, we intend to model these classes more realistically: for example, explicitly modelling a narrow jet that simultaneously carries kinetic luminosity and CRs. This raises a host of questions we have (again, intentionally) not tried to address here: for example, what happens if the injection is highly anisotropic? And can turbulence actually be driven by *physical* processes originating from an AGN? And what is the ratio of energy in radial momentum flux, thermal heating, CR injection, and turbulent stirring which comes from e.g. nuclear winds versus compact jets versus 'bubbles'? These and many more questions remain open and critical for progress in this field.

ACKNOWLEDGEMENTS

We thank Andrew Fabian for useful discussions and valuable comments. We also thank Eliot Quataert for conversations and collaboration. Support for PFH was provided by an Alfred P. Sloan Research Fellowship, NASA ATP Grant NNX14AH35G, and NSF Collaborative Research Grant #1411920 and CAREER grant #1455342. The Flatiron Institute is supported by the Simons Foundation. CAFG was supported by NSF through grants AST-1517491, AST-1715216, and CAREER award AST-1652522, by NASA through grant 17-ATP17-0067, by CXO through grant TM7-18007, and by a Cottrell Scholar Award from the Research Corporation for Science Advancement. DK was supported by NSF grant AST-1715101 and the Cottrell Scholar Award from the Research Corporation for Science Advancement. TKC was supported by NSF grant AST-1412153. VHR acknowledges support from UC-MEXUS and CONACyT through the postdoctoral fellowship. Numerical calculations were run on the Caltech compute cluster 'Wheeler', allocations from XSEDE TG-AST130039 and PRAC NSF.1713353 supported by the NSF, and NASA HEC SMD-16-7592.

REFERENCES

Ackermann M. et al., 2016, ApJ, 819, 149
Anderson M. E., Gaspari M., White S. D. M., Wang W., Dai X., 2015, MNRAS, 449, 3806
Anglés-Alcázar D., Faucher-Giguère C.-A., Kereš D., Hopkins P. F., Quataert E., Murray N., 2017, MNRAS, 470, 4698
Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, ApJ, 600, 681
Balogh M. L., Babul A., Voit G. M., McCarthy I. G., Jones L. R., Lewis G. F., Ebeling H., 2006, MNRAS, 366, 624
Banerjee N., Sharma P., 2014, MNRAS, 443, 687
Barai P., Viel M., Murante G., Gaspari M., Borgani S., 2014, MNRAS, 437, 1456
Bauer A., Springel V., 2012, MNRAS, 423, 2558
Beck A. M., Lesch H., Dolag K., Kotarba H., Geng A., Stasyszyn F. A., 2012, MNRAS, 422, 2152
Beck R., Brandenburg A., Moss D., Shukurov A., Sokoloff D., 1996, ARA&A, 34, 155
Begelman M. C., 2004, Coevolution of Black Holes and Galaxies. Cambridge University Press, Cambridge, p. 374
Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJS, 149, 289
Berezinsky V., Gazizov A., Grigorieva S., 2006, Phys. Rev. D, 74, 043005
Binney J., Cowie L. L., 1981, ApJ, 247, 464
Blanton M. R., Eisenstein D., Hogg D. W., Schlegel D. J., Brinkmann J., 2005, ApJ, 629, 143
Booth C. M., Schaye J., 2009, MNRAS, 398, 53
Bourne M. A., Sijacki D., 2017, MNRAS, 472, 4707
Bravi L., Gitti M., Brunetti G., 2016, preprint (arXiv:1603.00368)
Brighenti F., Mathews W. G., 2002, ApJ, 567, 130
Brüggen M., Scannapieco E., 2009, MNRAS, 398, 548
Butsky I. S., Quinn T. R., 2018, ApJ, 868, 108
Chan T. K., Kereš D., Hopkins P. F., Quataert E., Su K.-Y., Hayward C. C., Faucher-Giguère C.-A., 2019, MNRAS, 488, 3716
Choi E., Ostriker J. P., Naab T., Johansson P. H., 2012, ApJ, 754, 125
Choi E., Ostriker J. P., Naab T., Oser L., Moster B. P., 2015, MNRAS, 449, 4105
Cho J., Lazarian A., Honein A., Knaepen B., Kassinos S., Moin P., 2003, ApJ, 589, L77
Ciotti L., Ostriker J. P., 2001, ApJ, 551, 131
Ciotti L., Ostriker J. P., Proga D., 2009, ApJ, 699, 89
Ciotti L., Ostriker J. P., Proga D., 2010, ApJ, 717, 708
Conroy C., van Dokkum P. G., Kravtsov A., 2015, ApJ, 803, 77
Croton D. J. et al., 2006, MNRAS, 365, 11
Dekel A., Birnboim Y., 2006, MNRAS, 368, 2
Dekel A., Sari R., Ceverino D., 2009, ApJ, 703, 785
Di Matteo T., Springel V., Hernquist L., 2005, Nature, 433, 604
Dimonte G., Tipton R., 2006, Phys. Fluids, 18, 085101
Dubois Y., Pichon C., Devriendt J., Silk J., Haehnelt M., Kimm T., Slyz A., 2013, MNRAS, 428, 2885
Eisenreich M., Naab T., Choi E., Ostriker J. P., Emsellem E., 2017, MNRAS, 468, 751
Enßlin T., Pfrommer C., Miniati F., Subramanian K., 2011, A&A, 527, A99
Fabian A. C., 1999, MNRAS, 308, L39
Fabian A. C., 2012, ARA&A, 50, 455
Fabian A. C., Arnaud K. A., Bautz M. W., Tawara Y., 1994, ApJ, 436, L63
Fabian A. C., Voigt L. M., Morris R. G., 2002, MNRAS, 335, L71
Farber R., Ruszkowski M., Yang H.-Y. K., Zweibel E. G., 2018, ApJ, 856, 112
Faucher-Giguère C.-A., Quataert E., 2012, MNRAS, 425, 605
Federrath C., Roman-Duval J., Klessen R. S., Schmidt W., Mac Low M.-M., 2010, A&A, 512, A81
Fujita Y., Kimura S., Ohira Y., 2013, MNRAS, 432, 1434
Fujita Y., Matsumoto T., Wada K., 2004, ApJ, 612, L9
Fujita Y., Ohira Y., 2011, ApJ, 738, 182
Gaspari M., Brighenti F., Temi P., 2015, A&A, 579, A62
Gaspari M., Melioli C., Brighenti F., D'Ercole A., 2011, MNRAS, 411, 349
Gaspari M., Sądowski A., 2017, ApJ, 837, 149
Genel S. et al., 2014, MNRAS, 445, 175
Giacintucci S., Markevitch M., Venturi T., Clarke T. E., Cassano R., Mazzotta P., 2014, ApJ, 781, 9
Giodini S. et al., 2009, ApJ, 703, 982

Gonzalez A. H., Sivanandam S., Zabludoff A. I., Zaritsky D., 2013, ApJ, 778, 14
Graham J., Fabian A. C., Sanders J. S., Morris R. G., 2006, MNRAS, 368, 1369
Guo F., Oh S. P., 2008, MNRAS, 384, 251
Hernquist L., 1990, ApJ, 356, 359
Hickox R. C., Alexander D. M., 2018, ARA&A, 56, 625
Hitomi Collaboration, 2016, Nature, 535, 117
Hitomi Collaboration, 2018, PASJ, 70, 9
Hopkins P. F., 2015, MNRAS, 450, 53
Hopkins P. F., Elvis M., 2010, MNRAS, 401, 7
Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Martini P., Robertson B., Springel V., 2005, ApJ, 630, 705
Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2006a, ApJS, 163, 1
Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2006b, ApJS, 163, 1
Hopkins P. F., Hernquist L., Cox T. J., Kereš D., 2008, ApJS, 175, 356
Hopkins P. F., Hernquist L., Cox T. J., Robertson B., Krause E., 2007, ApJ, 669, 45
Hopkins P. F., Hernquist L., Cox T. J., Robertson B., Springel V., 2006c, ApJS, 163, 50
Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581
Hopkins P. F., Narayanan D., Murray N., 2013, MNRAS, 432, 2647
Hopkins P. F. et al., 2018a, MNRAS, 477, 1578
Hopkins P. F. et al., 2018b, MNRAS, 480, 800
Hudson D. S., Mittal R., Reiprich T. H., Nulsen P. E. J., Andernach H., Sarazin C. L., 2010, A&A, 513, A37
Humphrey P. J., Buote D. A., 2013, MNRAS, 436, 2879
Humphrey P. J., Buote D. A., Brighenti F., Flohic H. M. L. G., Gastaldello F., Mathews W. G., 2012, ApJ, 748, 11
Häring N., Rix H.-W., 2004, ApJ, 604, L89
Jacob S., Pakmor R., Simpson C. M., Springel V., Pfrommer C., 2018, MNRAS, 475, 570
Jacob S., Pfrommer C., 2017a, MNRAS, 467, 1449
Jacob S., Pfrommer C., 2017b, MNRAS, 467, 1478
Johansson P. H., Naab T., Burkert A., 2009, ApJ, 690, 802
Kaastra J. S., Mewe R., Nieuwenhuijzen H., 1996, in Yamashita K., Watanabe T., eds, UV and X-ray Spectroscopy of Astrophysical and Laboratory Plasmas, p. 411
Kauffmann G. et al., 2003, MNRAS, 341, 54
Kereš D., Katz N., Davé R., Fardal M., Weinberg D. H., 2009, MNRAS, 396, 2332
Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2
Kim D.-W., Fabbiano G., 2013, ApJ, 776, 116
Kim W.-T., Narayan R., 2003, ApJ, 596, L139
Kroupa P., 2002, Science, 295, 82
Leitherer C. et al., 1999, ApJS, 123, 3
Li Y., Bryan G. L., 2014, ApJ, 789, 54
Li Y., Bryan G. L., Ruszkowski M., Voit G. M., O'Shea B. W., Donahue M., 2015, ApJ, 811, 73
Li Y., Ruszkowski M., Bryan G. L., 2017, ApJ, 847, 106
Li Y.-P. et al., 2018, ApJ, 866, 70
Lodders K., 2003, ApJ, 591, 1220
Madgwick D. S., Somerville R., Lahav O., Ellis R., 2003, MNRAS, 343, 871
Martig M., Bournaud F., Teyssier R., Dekel A., 2009, ApJ, 707, 250
Martizzi D., Quataert E., Faucher-Giguere C.-A., Fielding D., 2019, MNRAS, 483, 2465
Mathews W. G., Faltenbacher A., Brighenti F., 2006, ApJ, 638, 659
McCarthy I. G. et al., 2010, MNRAS, 406, 822
McDonald M., Veilleux S., Mushotzky R., 2011, ApJ, 731, 33
McDonald M. et al., 2013, ApJ, 774, 23
McNamara B. R., Nulsen P. E. J., 2007, ARA&A, 45, 117
Meece G. R., Voit G. M., O'Shea B. W., 2017, ApJ, 841, 133
Mitchell N. L., McCarthy I. G., Bower R. G., Theuns T., Crain R. A., 2009, MNRAS, 395, 180
Mittal R., Hudson D. S., Reiprich T. H., Clarke T., 2009, A&A, 501, 835
Muratov A. L., Keres D., Faucher-Giguere C.-A., Hopkins P. F., Quataert E., Murray N., 2015, preprint (arXiv:1501.03155)
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Norman M. L., Bryan G. L., 1999, in Röser H.-J., Meisenheimer K., eds, Lecture Notes in Physics, Vol. 530, The Radio Galaxy Messier 87. Springer-Verlag, Berlin, p. 106
Ostriker J. P., Choi E., Ciotti L., Novak G. S., Proga D., 2010, ApJ, 722, 642
O'Dea C. P. et al., 2008, ApJ, 681, 1035
Parrish I. J., McCourt M., Quataert E., Sharma P., 2012, MNRAS, 419, L29
Parrish I. J., Quataert E., Sharma P., 2009, ApJ, 703, 96
Parrish I. J., Quataert E., Sharma P., 2010, ApJ, 712, L194
Paul S., Iapichino L., Miniati F., Bagchi J., Mannheim K., 2011, ApJ, 726, 17
Pellegrini S., Ciotti L., Negri A., Ostriker J. P., 2018, ApJ, 856, 115
Peres C. B., Fabian A. C., Edge A. C., Allen S. W., Johnstone R. M., White D. A., 1998, MNRAS, 298, 416
Peterson J. R., Fabian A. C., 2006, Phys. Rep., 427, 1
Pfrommer C., 2013, ApJ, 779, 10
Pfrommer C., Pakmor R., Schaal K., Simpson C. M., Springel V., 2017, MNRAS, 465, 4500
Pillepich A. et al., 2018, MNRAS, 473, 4077
Pozzetti L. et al., 2010, A&A, 523, A13
Price D. J., Federrath C., 2010, MNRAS, 406, 1659
Rafferty D. A., McNamara B. R., Nulsen P. E. J., 2008, ApJ, 687, 899
Rebusco P., Churazov E., Böhringer H., Forman W., 2005, MNRAS, 359, 1041
Rebusco P., Churazov E., Böhringer H., Forman W., 2006, MNRAS, 372, 1840
Reiprich T. H., Böhringer H., 2002, ApJ, 567, 716
Ressler S. M., Quataert E., Stone J. M., 2018, MNRAS, 478, 3544
Richings A. J., Faucher-Giguère C.-A., 2018a, MNRAS, 474, 3673
Richings A. J., Faucher-Giguère C.-A., 2018b, MNRAS, 478, 3100
Ricker P. M., Sarazin C. L., 2001, ApJ, 561, 621
Roettiger K., Burns J., Loken C., 1993, ApJ, 407, L53
Roettiger K., Loken C., Burns J. O., 1997, ApJS, 109, 307
Ruszkowski M., Oh S. P., 2010, ApJ, 713, 1332
Ruszkowski M., Oh S. P., 2011, MNRAS, 414, 1493
Ruszkowski M., Yang H.-Y. K., Reynolds C. S., 2017b, ApJ, 844, 13
Ruszkowski M., Yang H.-Y. K., Zweibel E., 2017a, ApJ, 834, 208
Sanderson A. J. R., O'Sullivan E., Ponman T. J., 2009, MNRAS, 395, 764
Sanderson A. J. R., O'Sullivan E., Ponman T. J., Gonzalez A. H., Sivanandam S., Zabludoff A. I., Zaritsky D., 2013, MNRAS, 429, 3288
Sanderson A. J. R., Ponman T. J., O'Sullivan E., 2006, MNRAS, 372, 1496
Scannapieco E., Brüggen M., 2008, ApJ, 686, 927
Schmidt W., Federrath C., Hupp M., Kern S., Niemeyer J. C., 2009, A&A, 494, 127
Schure K. M., Kosenko D., Kaastra J. S., Keppens R., Vink J., 2009, A&A, 508, 751
Sharma P., McCourt M., Quataert E., Parrish I. J., 2012, MNRAS, 420, 3174
Sharma P., Parrish I. J., Quataert E., 2010, ApJ, 720, 652
Shin M.-S., Ostriker J. P., Ciotti L., 2010, ApJ, 711, 268
Sijacki D., Springel V., Di Matteo T., Hernquist L., 2007, MNRAS, 380, 877
Silk J., Rees M. J., 1998, A&A, 331, L1
Smith B., O'Shea B. W., Voit G. M., Ventimiglia D., Skillman S. W., 2013, ApJ, 778, 152
Soker N., Sarazin C. L., 1990, ApJ, 348, 73
Springel V., 2000, MNRAS, 312, 859
Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776
Springel V., White S. D. M., 1999, MNRAS, 307, 162
Stanek R., Evrard A. E., Böhringer H., Schuecker P., Nord B., 2006, ApJ, 648, 956
Strong A. W., Moskalenko I. V., 1998, ApJ, 509, 212
Strong A. W., Porter T. A., Digel S. W., Jóhannesson G., Martin P., Moskalenko I. V., Murphy E. J., Orlando E., 2010, ApJ, 722, L58
Su K.-Y., Hopkins P. F., Hayward C. C., Faucher-Giguere C.-A., Keres D., Ma X., Robles V. H., 2017, MNRAS, 471, 144

Su K.-Y., Hopkins P. F., Hayward C. C., Ma X., Faucher-Giguère C.-A., Kereš D., Orr M. E., Robles V. H., 2019, MNRAS, 487, 4393 (Paper I)
Su Y., Buote D., Gastaldello F., Brighenti F., 2015, ApJ, 805, 104
Su Y., White R. E., III, Miller E. D., 2013, ApJ, 775, 89
Tamura T. et al., 2001, A&A, 365, L87
Trotta R., Jóhannesson G., Moskalenko I. V., Porter T. A., Ruiz de Austri R., Strong A. W., 2011, ApJ, 729, 106
Tucker W. H., Rosner R., 1983, ApJ, 267, 547
Uhlig M., Pfrommer C., Sharma M., Nath B. B., Enßlin T. A., Springel V., 2012, MNRAS, 423, 2374
Vazza F., Brunetti G., Gheller C., Brunino R., Brüggen M., 2011, A&A, 529, A17
Vernaleo J. C., Reynolds C. S., 2006, ApJ, 645, 83
Voigt L. M., Fabian A. C., 2004, MNRAS, 347, 1130
Voigt L. M., Schmidt R. W., Fabian A. C., Allen S. W., Johnstone R. M., 2002, MNRAS, 335, L7
Voit G. M., Meece G., Li Y., O'Shea B. W., Bryan G. L., Donahue M., 2017, ApJ, 845, 80
Weinberger R. et al., 2017a, preprint (arXiv:1710.04659)
Weinberger R. et al., 2017b, MNRAS, 465, 3291
Werner N. et al., 2013, ApJ, 767, 153
Wetzel A. R., Tinker J. L., Conroy C., 2012, MNRAS, 424, 232
Wiener J., Oh S. P., Guo F., 2013, MNRAS, 434, 2209
Wiener J., Zweibel E. G., 2019, MNRAS, 488, 280
Yang H.-Y. K., Reynolds C. S., 2016, ApJ, 818, 181
Yoon D., Yuan F., Gan Z.-M., Ostriker J. P., Li Y.-P., Ciotti L., 2018, ApJ, 864, 6
Zakamska N. L., Narayan R., 2003, ApJ, 582, 162
Zhuravleva I. et al., 2014, Nature, 515, 85
ZuHone J. A., Markevitch M., Brunetti G., Giacintucci S., 2013, ApJ, 762, 78
ZuHone J. A., Markevitch M., Zhuravleva I., 2016, ApJ, 817, 110
ZuHone J. A., Miller E. D., Bulbul E., Zhuravleva I., 2018, ApJ, 853, 180

APPENDIX A: EFFECTS OF MAGNETIC FIELDS, CONDUCTION, AND VISCOSITY ON TURBULENT 'STIRRING' MODELS

Given that turbulent stirring can (i) amplify magnetic fields, (ii) be damped by viscosity from the hot gas, and (iii) acts to mix hot and cold gas in a manner similar to physical conductivity, it is reasonable to ask what the impact of including or excluding explicit treatment of magnetic fields and physical (anisotropic) Braginskii conduction and viscosity in the hot gas might be. We explored these physics in Paper I in detail so only briefly note their effects here. Fig. A1 shows the SFRs of the 'Turb-core-1' run with and without explicit inclusion of these fluid microphysics in the simulations. Magnetic fields and conduction mildly suppress the SFR at the beginning of the 'Turb-core-1' run, and suppress the core baryonic mass by a factor of ∼2,

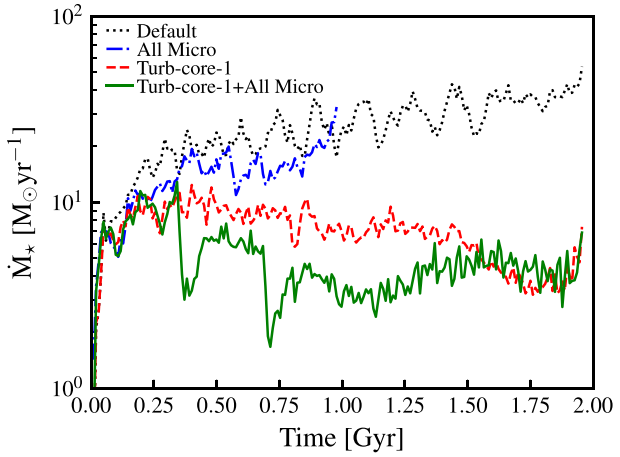

Figure A1. SFR (as Fig. 2) in our 'Default' and 'Turb-core-1' **m14** runs, comparing runs which treat the gas as pure hydrodynamic, to runs which include magnetic fields and fully anisotropic Spitzer–Braginskii conduction and viscosity following Su et al. (2017) ('All Micro'). Consistent with our study in Paper I, these additional microphysics (mostly conduction) suppress the SFRs by a factor ∼2, but do not qualitatively change any of our conclusions.

which is roughly consistent with their effect on the 'Default' run, but the systematic effects are small and get *smaller* as time goes on and the systems become more steady state. Because viscosity and conduction are strongly temperature dependent, their effects are even weaker in the smaller halo masses. Accordingly, the treatment of these physics does not substantially alter our conclusions.

APPENDIX B: EFFECTS OF TURBULENT 'STIRRING' ON THE METALLICITY DISTRIBUTION

Fig. B1 shows the effect of turbulent stirring on the metallicity profiles. The runs with turbulent stirring do not have flatter metallicity profiles, as has been suggested in the literature based on simulations employing an effective turbulent conduction treatment. Instead, the gas mass distribution, SFRs, and fraction of metals that is recycled have more significant effects. The total metallicity is primarily determined by the SFR, so the 'Default' (stellar feedback only) runs have the most metals at a given radius. However, there is also much more gas remaining within the same radius absent turbulent stirring, so the metallicities in the 'Turb' runs are not necessarily higher or lower than in the 'Default' runs.

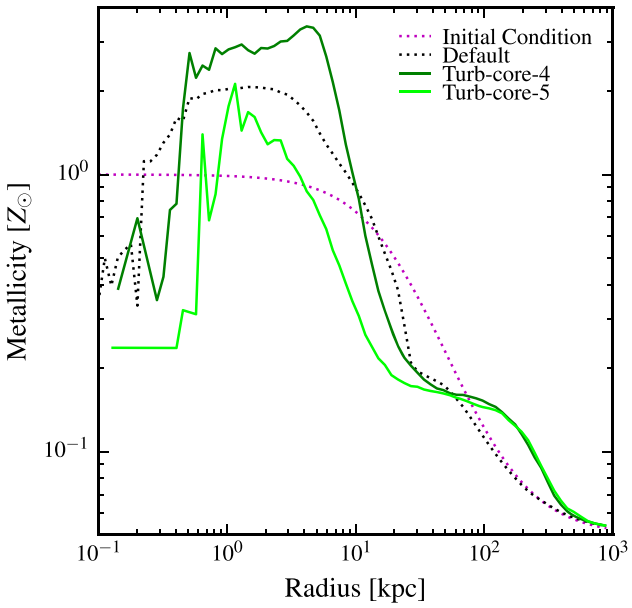

Figure B1. The metallicity profiles of different **m14** runs averaged over the last 100 Myr. The runs with turbulent stirring do not necessarily have flatter or steeper metallicity profiles. Instead, the gas mass distribution, the SFR, and the fraction of metals that is recycled have more significant effects.

APPENDIX C: MORPHOLOGIES OF THE MORE SUCCESSFUL RUNS

The face-on projected density and average temperature (within $\Delta Z = \pm 1$ kpc of the mid-plane) of the more successful runs ('Th-core-44', 'Turb-core-4' and 'CR-BH-43') are shown in Fig. C1. Consistent with the aforementioned density and temperature profiles, 'Th-core-44' and 'Turb-core-4' have suppressed densities up to a few tens of kpc, while 'CR-BH-43' has suppressed density only within 10 kpc. Thermal heating and CR injection both lead to a heated region, but the heated region in the former case extends to a larger radius.

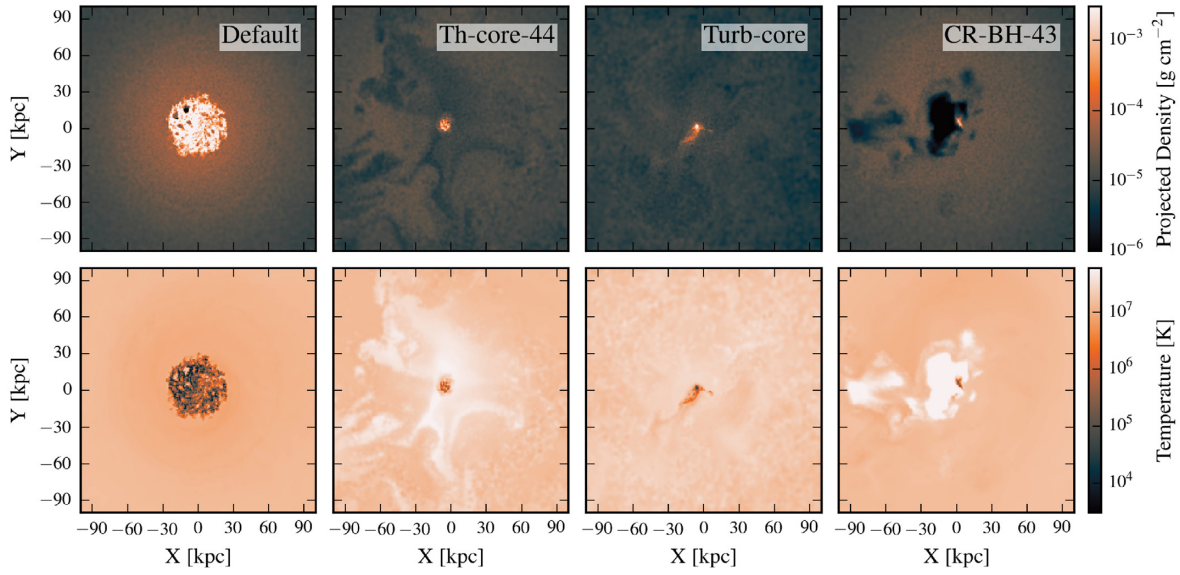

Figure C1. The face-on projected density and average temperature (between $\Delta Z = \pm 1$ kpc) of the more successful runs ('Th-core-44', 'Turb-core-4', and 'CR-BH-43'). Consistent with the density and temperature profiles in Fig. 3, 'Th-core-44' and 'Turb-core-4' have suppressed densities up to a few tens of kpc, while 'CR-BH-43' has suppressed density only within 10 kpc. Thermal heating and CR injection both lead to a heated region, but the heated region in the former case extends to a larger radius.

This paper has been typeset from a TEX/LATEX file prepared by the author.